

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:26:05 ; Search time 32 Seconds
(without alignments)
245.223 Million cell updates/sec

Title: US-10-768-874-2

Perfect score: 825

Sequence: 1 MSITQSFVLTIAIFGAASD.....NWGMKDGKHYFTVENCQE 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
5: /cgn2_6/ptodata/2/iaa/PCRUS COMB.pdp.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	825	100.0	152	3	US-09-036-355A-2
2	825	100.0	152	4	US-09-702-647A-2
3	610	73.9	109	3	US-09-036-355A-5
4	610	73.9	109	4	US-09-702-647A-5
5	130	15.8	22	3	US-09-036-355A-3
6	130	15.8	22	4	US-09-702-647A-3
7	88	10.7	712	4	US-09-059-584-51
8	87	10.5	15	3	US-09-036-355A-4
9	87	10.5	15	4	US-09-702-647A-4
10	84.5	10.2	1237	4	US-09-107-532A-4552
11	80.5	9.8	329	4	US-09-651-200-19
12	76	9.2	471	3	US-08-961-083-158
13	76	9.2	471	4	US-09-536-784-158
14	74	9.0	429	4	US-09-489-039A-12086
15	73.5	8.9	360	4	US-09-668-097A-32
16	72.5	8.8	911	1	US-08-487-890A-107
17	72.5	8.8	911	2	US-08-478-435-107
18	72.5	8.8	911	2	US-08-337-483-107
19	72.5	8.8	911	2	US-08-478-373-107
20	72.5	8.8	911	3	US-08-474-671-107
21	72.5	8.8	911	3	US-08-483-577A-107
22	72.5	8.8	911	3	US-08-897-438-107
23	72.5	8.8	911	4	US-08-637-654-107
24	72.5	8.8	911	4	US-08-649-518-107
25	72	8.7	616	4	US-09-608-790-1
26	72	8.7	943	4	US-09-397-885-5
27	72	8.7	943	4	US-09-969-362-5

28 71.5 8.7 363 4 US-09-134-001C-4219 Sequence 4219, Ap
29 70.5 8.5 367 3 US-09-041-889-30 Sequence 30, Appl
30 70.5 8.5 367 4 US-09-417-264-30 Sequence 30, Appl
31 70 8.5 330 4 US-09-296-840A-4 Sequence 4, Appl
32 70 8.5 1577 2 US-08-793-824-2 Sequence 2, Appl
33 69.5 8.4 325 4 US-09-651-200-20 Sequence 20, Appl
34 69.5 8.4 365 4 US-09-543-681A-6366 Sequence 6366, Ap
35 69.5 8.4 913 1 US-08-487-890A-5 Sequence 5, Appl
36 69.5 8.4 913 2 US-08-478-435-5 Sequence 5, Appl
37 69.5 8.4 913 2 US-08-337-483-5 Sequence 5, Appl
38 69.5 8.4 913 2 US-08-478-373-5 Sequence 5, Appl
39 69.5 8.4 913 3 US-08-474-671-5 Sequence 5, Appl
40 69.5 8.4 913 3 US-08-483-577A-5 Sequence 5, Appl
41 69.5 8.4 913 3 US-08-897-438-5 Sequence 5, Appl
42 69.5 8.4 913 4 US-08-637-654-5 Sequence 5, Appl
43 69.5 8.4 913 4 US-08-649-518-5 Sequence 5, Appl
44 69 8.4 923 4 US-09-397-885-1 Sequence 1, Appl
45 69 8.4 923 4 US-09-969-362-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-036-355A-2
; Sequence 2, Application US/09036355A
; Patent No. 6162785
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
; APPLICANT: Ribeiro, Jose M.C.
; APPLICANT: Swaim, Eddie W.
; APPLICANT: Swaim, Steven F.
; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
; TITLE OF INVENTION: SALIVARY GLAND OF THE BLACK FLY
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,355A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5721-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-355A-2

Query Match 100.0%; Score 825; DB 3; Length 152;
Best Local Similarity 100.0%; Pred No. 4,1e-92;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSITQSFVLTIAIFGAASDNPFIADKRCIVISDGLVWHERKPGQFFYYVIMPKGTEY 60
Db 1 MSITQSFVLTIAIFGAASDNPFIADKRCIVISDGLVWHERKPGQFFYYVIMPKGTEY 60

QY 61 DDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAVREMDHFKFTADGTGKYDIS 120
 Db 61 DDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAVREMDHFKFTADGTGKYDIS 120
 QY 121 SKANGHPRSRGNKGNWGMKDGKHFTVENCQE 152
 Db 121 SKANGHPRSRGNKGNWGMKDGKHFTVENCQE 152

RESULT 2
 US-09-702-647A-2
 ; Sequence 2, Application US/09702647A
 ; Patent No. 6500420
 ; GENERAL INFORMATION:
 ; APPLICANT: Cupp, Mary S.
 ; APPLICANT: Ribeiro, Jose M.C.
 ; APPLICANT: Cupp, Eddie W.
 ; APPLICANT: Swaim, Steven F.
 ; TITLE OF INVENTION: Recombinant Vasoactive Protein from
 ; TITLE OF INVENTION: Salivary Gland of the Black Fly
 ; FILE REFERENCE: 35721/205407
 ; CURRENT APPLICATION NUMBER: US/09/702,647A
 ; CURRENT FILING DATE: 2000-10-31
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: Fast-Seq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 152
 ; TYPE: PRT
 ; ORGANISM: Simulium vittatum
 US-09-702-647A-2

Query Match 100.0%; Score 825; DB 4; Length 152;
 Best Local Similarity 100.0%; Pred. No. 4.1e-92; Indels 0; Gaps 0;
 Matches 152; Conservative 0; Mismatches 0;
 QY 1 MSITQSFVLTALIFGAASDNPIADKICIVISDGLVWHERKPGQEFFYVYVIMIPKGTGY 60
 Db 1 MSITQSFVLTALIFGAASDNPIADKICIVISDGLVWHERKPGQEFFYVYVIMIPKGTGY 60
 QY 61 DDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAVREMDHFKFTADGTGKYDIS 120
 Db 61 DDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAVREMDHFKFTADGTGKYDIS 120
 QY 121 SKANGHPRSRGNKGNWGMKDGKHFTVENCQE 152
 Db 121 SKANGHPRSRGNKGNWGMKDGKHFTVENCQE 152

RESULT 3
 US-09-036-355A-5
 ; Sequence 5, Application US/09036355A
 ; Patent No. 6162785
 ; GENERAL INFORMATION:
 ; APPLICANT: Cupp, Mary S.
 ; APPLICANT: Ribeiro, Jose M.C.
 ; APPLICANT: Cupp, Eddie W.
 ; APPLICANT: Swaim, Steven F.
 ; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
 ; TITLE OF INVENTION: SALIVARY GLAND OF THE BLACK FLY
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
 ; STREET: 3605 Glenwood Ave. Suite 310
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA
 ; ZIP: 27622
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/036,355A
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Spruill, W. Murray
 ; REGISTRATION NUMBER: 32,943
 ; REFERENCE/DOCKET NUMBER: 5721-4
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919 420 2202
 ; TELEFAX: 919 881 3175
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 109 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-09-036-355A-5

Query Match 73.9%; Score 610; DB 3; Length 109;
 Best Local Similarity 100.0%; Pred. No. 3e-66;
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 42 KPQGEFFYVYVIMIPKGTGYDDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAV 101
 Db 1 KPQGEFFYVYVIMIPKGTGYDDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAV 60
 QY 102 REMDHFKFTADGTGKYDISKANGHPRSRGNKGNWGMKDGKHFTVENC 150
 Db 61 REMDHFKFTADGTGKYDISKANGHPRSRGNKGNWGMKDGKHFTVENC 109

RESULT 4
 US-09-702-647A-5
 ; Sequence 5, Application US/09702647A
 ; Patent No. 6500420
 ; GENERAL INFORMATION:
 ; APPLICANT: Cupp, Mary S.
 ; APPLICANT: Ribeiro, Jose M.C.
 ; APPLICANT: Cupp, Eddie W.
 ; APPLICANT: Swaim, Steven F.
 ; TITLE OF INVENTION: Recombinant Vasoactive Protein from
 ; TITLE OF INVENTION: Salivary Gland of the Black Fly
 ; FILE REFERENCE: 35721/205407
 ; CURRENT APPLICATION NUMBER: US/09/702,647A
 ; CURRENT FILING DATE: 2000-10-31
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 109
 ; TYPE: PRT
 ; ORGANISM: Simulium vittatum
 US-09-702-647A-5

Query Match 73.9%; Score 610; DB 4; Length 109;
 Best Local Similarity 100.0%; Pred. No. 3e-66;
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 42 KPQGEFFYVYVIMIPKGTGYDDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAV 101
 Db 1 KPQGEFFYVYVIMIPKGTGYDDQWILESGDHYKLNKPSGRYLVTGTFDYFLTAGAAV 60
 QY 102 REMDHFKFTADGTGKYDISKANGHPRSRGNKGNWGMKDGKHFTVENC 150
 Db 61 REMDHFKFTADGTGKYDISKANGHPRSRGNKGNWGMKDGKHFTVENC 109

RESULT 5
 US-09-036-355A-3
 ; Sequence 3, Application US/09036355A
 ; Patent No. 6162785
 ; GENERAL INFORMATION:

APPLICANT: Cupp, Mary S.
APPLICANT: Ribeiro, Jose M.C.
APPLICANT: Cupp, Eddie W.
APPLICANT: Swaim, Steven F.
TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
TITLE OF INVENTION: SALIVARY GLAND OF THE BLACK FLY
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
STREET: 3605 Glenwood Ave. Suite 310
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27622
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,355A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: 5721-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2202
TELEFAX: 919 881 3175
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-036-355A-3

Query Match 15.8%; Score 130; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 4.8e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 GKWGVKDGKHYFTVENCQ 152
Db 1 GKWGVKDGKHYFTVENCQ 22

RESULT 6
US-09-702-647A-3
Sequence 3, Application US/09702647A
Patent No. 650420
GENERAL INFORMATION:
APPLICANT: Cupp, Mary S.
APPLICANT: Ribeiro, Jose M.C.
APPLICANT: Cupp, Eddie W.
APPLICANT: Swaim, Steven F.
TITLE OF INVENTION: Recombinant Vasoactive Protein from
TITLE OF INVENTION: Salivary Gland of the Black Fly
FILE REFERENCE: 35721/205407
CURRENT APPLICATION NUMBER: US/09/702,647A
CURRENT FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 22
TYPE: PRT
ORGANISM: Simulium vittatum
US-09-702-647A-3

Query Match 15.8%; Score 130; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 4.8e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 GKWGVKDGKHYFTVENCQ 152
Db 1 GKWGVKDGKHYFTVENCQ 22

RESULT 7
US-09-059-584-51
Sequence 51, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loosmore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-51

Query Match 10.7%; Score 88; DB 4; Length 712;
Best Local Similarity 27.8%; Pred. No. 0.09;
Matches 37; Conservative 10; Mismatches 44; Indels 42; Gaps 6;

Qy 38 MHERKQGPYVYVIMPKGTEDYDQWILESVGGGHYKL-----KMKFSGR 84
Db 276 MTDKKGQSFSSP-----ERRAGDEYSAMSSHEYPSSLTDDKPKPDNY 318
Qy 85 YLYVG-----TFDY---FLTAGAAVREMDHFKFTADGTGKYDISSKANGHPRGRKNGV 136
Db 319 NDEYGHSSSEPTVDFSKSLTGLFSLNQDHHKGVTKTKYDINATHGN-RFRGSATAI 377
Qy 137 MKDGE---KHYPT 146
Db 378 NKDNESKAKHPFT 390

RESULT 8

US-09-036-355A-4
; Sequence 4, Application US/09036355A
; Patent No. 6162785
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
; APPLICANT: Ribeiro, Jose M.C.
; APPLICANT: Cupp, Eddie W.
; APPLICANT: Swaim, Steven F.
; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
; TITLE OF INVENTION: SALIVARY GLAND OF THE BLACK FLY
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,355A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5721-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-036-355A-4

Query Match 10.5%; Score 87; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00045; Indels 0;
Matches 15; Conservative 0; Mismatches 0; Gaps 0;

QY 42 KPGQEFFYYVMIPK 56
Db 1 KPGQEFFYYVMIPK 15

RESULT 9
US-09-702-647A-4
; Sequence 4, Application US/09702647A
; Patent No. 6500420
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
; APPLICANT: Ribeiro, Jose M.C.
; APPLICANT: Cupp, Eddie W.
; APPLICANT: Swaim, Steven F.
; TITLE OF INVENTION: Recombinant Vasoactive Protein from
; TITLE OF INVENTION: Salivary Gland of the Black Fly
; FILE REFERENCE: 35721/205407
; CURRENT APPLICATION NUMBER: US/09/702, 647A
; CURRENT FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Simulium vittatum
US-09-702-647A-4

Query Match 10.5%; Score 87; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00045; Indels 0;
Matches 15; Conservative 0; Mismatches 0; Gaps 0;

QY 42 KPGQEFFYYVMIPK 56
Db 1 KPGQEFFYYVMIPK 15

RESULT 10
US-09-107-532A-4552
; Sequence 4552, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4552:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1297
; SEQUENCE DESCRIPTION: SEQ ID NO: 4552:
US-09-107-532A-4552

Query Match 10.2%; Score 84.5; DB 4; Length 1297;
Best Local Similarity 28.8%; Pred. No. 0.57; Indels 39; Gaps 8;
Matches 36; Conservative 14; Mismatches 36; Indels 39; Gaps 8;

QY 30 VISGDLVWHE-----RKPGQEFFYYV-YMI-----PKGTFFDDOR--WIL-----ES 69
Db 920 VIVDGEKVIHPIGEPDPSFVYKDSSEETNYAYKIESQNPGEFKPEPDNDTWILLYDNA 979
QY 70 VGGDHYKLKFKSGRYL-----VGTDFYELTAGAAVEMDHFKFTADGT 114
Db 980 IGTTLKQGTGENWQYLYTTTDEEDRLWVAFTAYNTFDY---REIRVRKLDHEGHTLQGA 1036

Matches 30; Conservative 19; Mismatches 48; Indels 22; Gaps 6;
QY 34 GDLVHRRKPGQEPFYVYVIMPKCHTEYDDORWILESVGGDHYKLNKFSGRYL 86
DB 156 GAMQONELWDPAISAYFLKSDGT-YANQEW--QKVGKWWYFFKKWGYMARNEWQG--- 209
QY 87 VYGTDFYELTAGAAVREMDHFKFTADGTGKYDISSKANGHPRSRCKNWGMKDGSKHYF 145
DB 210 ----NYILTGSAA---NATDEVIMDGTFTYIPAASGELKEKDLNVGW-VHRDGRKYFF 259

RESULT 14

US-09-489-039A-12086
; Sequence 12086, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12086
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12086

Query Match 9.0%; Score 74; DB 4; Length 429;

Best Local Similarity 31.7%; Pred. No. 2.2;
Matches 32; Conservative 16; Mismatches 37; Indels 16; Gaps 6;
QY 2 SITQSFVLTALFAGASDNPIADRKCIVISDGLVHMERKPGQEPFYVYVIMPKGTRYD 61
DB 169 SVSSQF--LTALLMTA---PLAQDTVIVIKGDLV---SKFYIDITLHL-MKTFGVAVD 218
QY 62 DORWILESVGGDHYKLNKFSGRYLVG---TFDYFLTAGA 99
DB 219 NQSYORFVVRG---KQYQSPGDYLVREGDASSYFLAAGA 256

RESULT 15

US-09-668-097A-32
; Sequence 32, Application US/09668097A
; Patent No. 6673988
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Edgar B.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Klein, Ted
; APPLICANT: Shen, Jennie
; TITLE OF INVENTION: Plant Lipases
; FILE REFERENCE: BB1401 US NA
; CURRENT APPLICATION NUMBER: US/09/668,097A
; CURRENT FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/157309
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-668-097A-32

Query Match 8.9%; Score 73.5; DB 4; Length 360;

Best Local Similarity 22.9%; Pred. No. 1.9;
Matches 38; Conservative 25; Mismatches 52; Indels 51; Gaps 8;
QY 3 ITQSFVLT-----TLAIFGAAS-----DNPIADRKCIVISDG--DLVMHER 41

Db 153 LSKSLFILGERFGNDYNAAIFGKSLDEVYTVVPHI INKVTSGVETLIGLGAVDVVVPGV 212
QY 42 KPGQEPFY--VYVIMPKGTEYD-----DORWILESVGGDHYKLNKFSGRY 85
DB 213 LP1GCFPLYLTLYGSSNQSDYDGDCLRRFNDLSGYHNRLLRQIGIG----RLRSKYAGVR 268
QY 86 LVYGTDFYELTAGAAVREMDHFKFT-----ADGTGKYDISSKA 123
DB 269 LMYG--DFYTQVAENVRSPRSFGLDYGLTVCCGASGGQSYNNYNNKA 312

Search completed: September 9, 2004, 15:32:04
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:22:35 ; Search time 40 seconds
(without alignments)
365.528 Million cell updates/sec

Title: US-10-768-874-2

Perfect score: 825

Sequence: 1 MSITQSPFVLTALFQAASD.....NWGVMDKGEKHYFTVENCQE 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.5	10.1	474	2 S76904	hypothetical prote
2	78	9.5	448	2 C98122	choline binding pr
3	78	9.5	1143	2 S46122	SNF2 protein homol
4	77	9.3	174	2 S07146	gamma-s-crystallin
5	77	9.3	501	2 F98050	lysozyme (EC 3.2.1
6	77	9.3	557	2 D97210	extracellular neut
7	76.5	9.3	332	2 B90522	hypothetical prote
8	76.5	9.3	411	2 JC7869	cell wall-lytic be
9	76.5	9.3	768	2 B97725	outer membrane pro
10	75.5	9.2	245	1 S74326	sensory transducti
11	75.5	9.2	444	2 B64119	glucose-1-phosphat
12	75	9.1	448	2 C95257	choline binding pr
13	75	9.1	673	2 T48701	hypothetical prote
14	75	9.1	1013	1 S00801	Na+/K+-exchanging
15	75	9.1	1013	2 C24639	Na+/K+-exchanging
16	73.5	8.9	613	2 A99552	oligonucleotidase
17	73	8.8	277	2 S17600	myelin proteolipid
18	73	8.8	277	2 I51270	myelin proteolipid
19	73	8.8	870	2 S27514	mosquitocidal toxi
20	72.5	8.8	164	2 G96632	hypothetical prote
21	72.5	8.8	581	2 G90181	hypothetical prote
22	72.5	8.8	768	2 D71726	outer membrane pro
23	72.5	8.8	911	2 S70911	transferrin-bindin
24	72	8.7	616	2 A55796	scarlin precursor
25	71.5	8.7	238	2 AG2396	two-component resp
26	71.5	8.7	303	2 F86158	alternative splic
27	71.5	8.7	640	2 B32355	hypothetical prote
28	71.5	8.7	656	2 T03473	acetate-CoA ligase
29	71	8.6	434	2 B81716	hypothetical prote

30	71	8.6	1310	2 AD1380	glycosidase homolo
31	70.5	8.5	367	1 MMECP	outer membrane por
32	70.5	8.5	448	2 AI0442	probable exported
33	70.5	8.5	516	2 E70035	levanase homolog y
34	70.5	8.5	559	2 H83758	alpha,alpha-phosph
35	70.5	8.5	722	2 H96986	endo-1,4-beta gluc
36	70.5	8.5	1139	2 S28277	hypothetical prote
37	70.5	8.5	1178	2 AD1845	5-methyltetrahydro
38	70	8.5	323	2 AD1963	hypothetical prote
39	70	8.5	330	1 JS0102	site-specific DNA-
40	70	8.5	435	2 AE3425	xylose isomerase (
41	70	8.5	620	2 JC4925	xmni methyltransfe
42	70	8.5	1577	2 T30858	glycosyltransferas
43	69.5	8.4	162	2 A96884	hypothetical prote
44	69.5	8.4	604	2 S69793	glutamine-fructose
45	69.5	8.4	656	2 H84206	acetyl-CoA synthet

ALIGNMENTS

RESULT 1

S76904

hypothetical protein - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S76904

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

S:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76904

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-474 <KAN>

A:Cross-references: EMBL:D90917; GB:AB001339; MID:g1653836; PID:BAA18816.1; PID:g16539

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synecocystis hypothetical protein slr0459

Query Match 10.1%; Score 83.5; DB 2; Length 474;

Best Local Similarity 27.6%; Pred.No. 2;

Matches 32; Conservative 11; Mismatches 48; Indels 25; Gaps 6;

QY 42 KPQGEFFYYVMIPKGTEDDQR--W-----ILESCVGDHYKLNKFKSGRYLVYGTDFYF 94

Db 356 QPGQTISYSLSKPLKGGEPASVTLVWERLVQLDLSNGNQYDLGESFQAQPL--SNLDLF 413

QY 95 LT--AGAA-----VREMDHFKFTADGTGKYDISSKANG-----HPSRSGKNW 134

Db 414 ITDQGGATPVCCSSQSKVDNVEHFLCPIASTGKYTIQVHKHQGGTAPRPENVALSW 469

RESULT 2

C98122

choline binding protein D [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: C98122

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; F

Y, P.; Leblanc, D.J.; Lee, L.N.; Lerkowitz, E.J.; Lu, J.; Matsushima, F.; McAhren, S.; M

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: C98122

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-448 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAL00808.1; PID:g15459711; GSPDB:GN00174

C:Genetics:
A:Gene: cbpb

Query Match 9.5%; Score 78; DB 2; Length 448;
Best Local Similarity 23.8%; Pred. No. 6.5;
Matches 41; Conservative 19; Mismatches 38; Indels 74; Gaps 11;
Qy 43 PGQFFPYVYVMPKTEYDDQW-----ILESVDGHHY- 75
Db 217 PGRVHVQDIL-----EKDGYKLSYTAINGSYRVQLEAVKNPLGNSVLSSTGGTHYF 271
Qy 76 KLNKFSGRVLYVGT-FDYELTAGAA-----VREMDHFK---FTA-DGFGKY----- 117
Db 272 KIRSAIKTEPLVSAIVDIYTPGKVKHQDILEDKYKWLSTYTAINGRRRIQLEGVTS 331
Qy 118 -----DISS-----KANG---HPRSRG-KNKGVMKDGKHYF 145
Db 332 QNYQNSGNISSYSGNSSTVGNKINGSWYHFKSNGSKSTGWLKDGSSWY 383

RESULT 3

S46122
SNF2 protein homolog YBR245C - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR1633
C:Species: Saccharomyces cerevisiae
C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 08-Sep-2002
C:Accession: S46122
R:Aljinovic, G.; Pohl, F.M.; Pohl, T.M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45906
A:Accession: S46122
A:Molecule type: DNA
A:Residues: 1-1143 <ALJ>
A:Cross-references: EMBL:Z36114; MIPS:YBR245C
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:ISWI
A:Cross-references: SGD:S0000449
A:Map position: 2R
C:Superfamily: chromatin remodelling complex ATPase chain ISWI; bromodomain homology
C:Keywords: nucleus; transmembrane protein
F:305-321/Domain: transmembrane #status predicted <TMM>

Query Match 9.5%; Score 78; DB 2; Length 1143;
Best Local Similarity 23.3%; Pred. No. 19;
Matches 37; Conservative 34; Mismatches 56; Indels 32; Gaps 9;
Qy 5 QSEFFVLTALFGAASNDPIADRK---CIVISDGLVMHMKPGQBF-----PYVYVMPK 56
Db 49 ENLFLKILALF-----LQEKKYDC--ISSGSIM---TASEEYLENLEKPFQVGLPH 95
Qy 57 GFEYDDQRMILSVGDHVKLK---NKFSGRVLYVGTFTDYFLTAGAAVREMDHFKFTADG 113
Db 96 DPESNKKRYLLKDANGKFDLEGTWTFKFEHLLSLGLFKHTIESKAA--KDPKFRQVLDV 153
Qy 114 TQYDILSSKANGHPSRGKNGVMKDGKHYFVENCQE 152
Db 154 LEE---NKANG--KKGKHQDVRRRKTEHEDAEILLKE 186

RESULT 4

S07146
gamma-s-crystallin - common carp
N:Alternate names: beta-s-crystallin
C:Species: Cyprinus carpio (common carp)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999
C:Accession: S07146
R:Chang, T.; Chang, W.C.
Biochim. Biophys. Acta 910, 89-92, 1987
A:Title: Cloning and sequencing of a carp beta-s-crystallin cDNA.
A:Reference number: S07146; MUID:88000706; PMID:3307924
A:Accession: S07146
A:Molecule type: mRNA

A:Residues: 1-174 <CHA>
A:Cross-references: EMBL:M6834; NID:g213051; PIDN:AAA49206.1; PID:g213052
C:Superfamily: beta-crystallin
C:Keywords: duplication

Query Match 9.3%; Score 77; DB 2; Length 174;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 36; Conservative 18; Mismatches 41; Indels 40; Gaps 9;
Qy 29 IVISDGLVMHMKPGQFFPYVYVMPKTEYDDQW-----ILESVDGHHYK 77
Db 36 IRVESGAWVIYER---PNEFGYQVLTGRGEYDQRMWGLNDRCLSCCKMIHFVSGSEYKI 92
Qy 78 ----KNKFSGRVLYV-----GTFDFLTAGAAVREMDHFKFTADGFGKYDIDSKANGH 126
Db 93 QLYDKGDTGQ--VIRESTEDCFVVDREPT-----REVHSCK-VLDGIWIF-----YEH 138
Qy 127 PRSRGKNWGMKGE 141
Db 139 PNYRGRQY-LLEKGE 152

RESULT 5

F98050
Lysozyme (EC 3.2.1.17) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: F98050
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F98050
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-501 <KUP>
A:Cross-references: GB:AB007317; PIDN:AAL00235.1; PID:g15459086; GSPDB:GN00174
C:Genetics:
A:Gene: lytC
C:Keywords: glycosidase; hydrolase

Query Match 9.3%; Score 77; DB 2; Length 501;
Best Local Similarity 25.0%; Pred. No. 9.3;
Matches 31; Conservative 18; Mismatches 43; Indels 32; Gaps 8;
Qy 34 GDLVMHMKPGQFFPYVYVMPKTEYDDQRMILSVGDHVKLK-----NKFSGRYL 86
Db 186 GAMIQNEWLYDPAYSAYFLKSDGT-YANQEW--QKVGKWIYFKKGTWGRNARWQGNV- 241
Qy 87 VGTDFYFLTAGAAVRE---MD--HFKFTADGFGKYDIDSKANGHPSRGKNGVMKGE 141
Db 242 -----YLTGSGMATDEVIDMGARYIFAASELK-----EKKOLNVGM-VHRDCK 285
Qy 142 KHYF 145
Db 286 RYFF 289

RESULT 6

D97210
extracellular neutral metalloprotease, NPPE [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 02-Jun-2003
C:Accession: D97210
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97210

A;Accession: JC7869
A;Molecule type: DNA
A;Residues: 1-411 <ASA>
A;Cross-references: DDBJ:AB078775
A;Accession: PC7199
A:Molecule type: protein
A;Residues: 28-42;95-111;152-170 <AS2>
C;Comment: This protein that is a fungal cell wall-degrading enzyme, hydrolyze
C;Genetics:
A;Gene: bgIM

Query Match 9.3%; Score 76.5; DB 2; Length 411;
Best Local Similarity 23.0%; Pred. No. 8.3;
Matches 35; Conservative 23; Mismatches 59; Indels 35; Gaps 6

QY 7 FFVLTLAIFG-----AASDNPIDRKC-----VISGDVLWHERKPGQBFP 48
| | | | : : : :
Db 227 FILNLAVGNWPSPNNSTPFPSQLVDYVRVTQDTGASNVISDGIYTIASKASKGVMD 286
| | | | : : : :
QY 49 YYYVMIPKG-----TRY--DDQRWILESGDHVKLNKFSGRVLNVGTDPDYFLTAGAA 100
| | | | : : : :
Db 287 VDVYSTARGAKIQOWTVANNQFRVESTGGYIKLTAHSGKLVDPSSS--TSTGVQ 344
| | | | : : : :
QY 101 VREMD-----HPKFTADGTGKYDISSKANG 125
| | | | : : : :
Db 345 LQQWDNDNGSNAQRWKIVDVGGYKLVSKVSQ 376
| | | | : : : :

RESULT 9
B97725
outer membrane protein omp1 [imported] - Rickettsia conorii (strain Maliss)
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C;Accession: B97725
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii
A;Reference number: A97700; PMID:21442074; PMID:11557893
A;Accession: B97725
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 1-768 <KUR>
A;Cross-references: GB:AE006914; PIDN:AAL02740.1; PID:g15619252; GSPDB:GN00173
C;Genetics:
A;Gene: omp1
C;Superfamily: protective surface antigen D-15

Query Match 9.3%; Score 76.5; DB 2; Length 768;
Best Local Similarity 23.1%; Pred. No. 17;
Matches 30; Conservative 24; Mismatches 47; Indels 29; Gaps 7

QY 34 GDLMVHERKPGQFPYYVYMIPKTEYDDQRWILESGGDHYKLKNFSGRY---LVYGT 90
| | | | : : : :
Db 563 GHITITYQTDNKIYFKNGLYVSGQGEP-----AGVGNKKYIKHEVDYGKYSFIHNK 615
| | | | : : : :
QY 91 FDFELTA-----GAAYREMDDHKFTADGTGKYDISSKANG--HPRSRGNKWGMKD 139
| | | | : : : :
Db 616 LTUKLASGGNITGLGKIRISDRFML-----GDYSLRGFGANGVGPREKVTFEGL--G 668
| | | | : : : :
QY 140 GEKHYPFVE 148
| | | | : : : :
Db 669 GERIYKFSTE 678
| | | | : : : :

RESULT 10
S74926
sensory transduction system regulatory protein sl10649 - Synechocystis sp. (str.
N;Alternate names: protein sl10649
C;Species: Synechocystis sp.
A;Variety: FCC 6803
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S74926
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miya-

O, K.; Okumura, S.; Shimpo, S.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S. *DNA Res.* 3, 109-136, 1996

A;Reference number: S74322; MUUD:97061201; PMID:8905231
A;Accession: S74326
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-245 <KAN>
A;Cross-references: EMBL:D30902; GB:AB001339; PIDN:BAA16966.1; PID:g165204
A;Note: the nucleotide sequence was submitted to the EMBL Data Library. June 1996

A;Accession:the nucleotide sequence was submitted to the EMBL Data Library; June 1996

C;Genetics:

A;Gene: phoP

C;Superfamily: ompR protein; response regulator homology

C;Keywords: phosphoprotein; transcription regulation

F;1-125/Domain: response regulator homology <RR>

E;63/Binding site: phosphate (Acn) (covalent) status predicted

Query Match 9.2%; Score 75.5; DB 1; Length 245;
Best local Similarity 30.7%; Pred. No. 5.8;
Matches 27: Conservative 20: Indels 31: Gaps 4:
F;62/Binding site: phosphate (asp) (covalent) #status predicted

	Matches	21; Conservative	10; Mismatches	20; Indels	31; Gaps	4;
QY	28	CIVISDG-----DLVWHE-----RKPGQEFYVYVMPKGTG	59			
Dp	38	CIVNDEGTGIFQFOVPLVLDLMLPKLDGLEVCTRIPOOPGSKDPLMLTKAGE	97			

Dc		60	YDQRWLESVGGDHVKLNKPFSGRYLV	87
Dc		98	ID--RIIGLSTGADDYLWK-PFGPRELV	122
Dd				
Qy				

RESULT 11
B64119

glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) - Haemophilus influenzae (strain
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Jun-1999
C;Accession: B64119

R; Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.; Guyonay, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
A:Reference number: A64000; MUID: 95350630; PMID: 7542800
A:Accession: B64119
A:Accession multiple: none
A:Accession multiple comment: none

A: Molecule type: DNA
A: Residues: 1-444 <TIGR>
A: Cross-references: GB:U32815; GB:I42023; NID:g1574818; PIDN:AAC23006.1; PID:g1574822; TIGR:19222
C: Superfamily: glucose-1-phosphate adenylyltransferase
A: Status: nucleic acid sequence not shown; translation not shown

C;Superfamily: glucose-1-phosphate acetyltransferase
C;Keywords: glycogen/starch biosynthesis; nucleotidyltransferase
F;45/Binding site: fructose-1,6-bisphosphate (lys) #status predicted
F;120/Binding site: substrate (tyr) #status predicted
F;202/Binding site: substrate (lys) #status predicted

Query Match 9.2% Score 75.5; DB 2; Length 444;
Best Local Similarity 26.7%; Pred. No. 11;
Matches 27; Conservative 16; Mismatches 47; Indels 11; Gaps 3;
r,2wz/running site: substrate (nfs) #status predicted

[illegible]

Qy	102	REMDHFKFTADGT-GKYDISSKANGHPRSRGKNWGVNMGDE	141
		: :	
Db	159	MLMDHVNSGAKCTVGCIEV-----PRSEAHEFGWAVNE	192

RESULT 12
C95257

choline binding protein D [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text change 03-Aug-2001

C/Accession: C95257
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heinson, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle, nson, T.; Hickey, E.K.; Hoot, I.E.

Science 293, 498-506, 2001
A:Authors: Lotz, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a Virulent isolate of *Streptococcus pneumoniae*.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95257
A:Status: preliminary
A:Molecule type: DNA

A;Residue: 1-448 <KUR>
A;Residue type: SAR
A;Cross-references: GB:AE005673; PID:g14973713; GSPDB:GN00164; TIGR:SP2201
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP2201

Query Match 9.1%; Score 75; DB 2; Length 448;
 Best Local Similarity 23.8%; Pred. No. 13;
 Matches 41; Conservative 18; Mismatches 39; Indels 74; Gaps 11;

Qy 43 PGOEFFYVWPKGTYYDQRW-----LLESGGDHY- 75
||::| | : | | : ||
Db 217 PGEKVHVDQIL-----EKDGKWLSTAYNGSYRYVQLNAVKNPLGNSVLSSGGTHYF 271

Qy	76	KLNKFGSRVLYGCT-FDYELTAGAA-----VREMDHFK-----FTA-DGTCKY-----	117
		:	
		:	
		:	
Db	272	KTKSAIKTEPLVSATVIDYYPGKRVHYDQILEKGKWLSTAYTANGSRRRIQLGEVTSS	331

```

Qy 118 -----DISS-----KANG---HPRSRC-KNVGVNMGDKGKHYF 145
Db 332 QNYQNGSNTSSYGSHSSITVGWKKINGSWYHFKSNGSKSTGLWLGKSSWY 383

```

RESULT 13
T48701
hypothetical protein 1A9.50 [imported] - Neurospora crassa

C:Species: Neurospora crassa
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48701
R:Schulte, U.; Agn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura
B., and
Schulte, U. 2000. The genome of Neurospora crassa: a model for eukaryotic
genome organization. *Genome Res.* 10: 1301-1313.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24541
A;Accession: T48701
A;Status: preliminary
A;Molecular structure: DVA

A;Molecule type: DNA
A;Residues: 1-673 <SCH>
A;Cross-references: EMBL:AI35317; GSPDB:GN00112; NCSP:1A9.50
A;Experimental source: cosmid contig 1A9; strain 74
C;Genetics.

C; Genetics:
A; Gene: NCSP:IA9.50
A; Map position: 2
A; Introns: 168/1; 316/2; 382/1; 626/3

Query Match 9.1%; Score 75; DB 2; Length 673;
 Best Local Similarity 29.4%; Pred. No. 21;
 Matches 32; Conservative 9; Mismatches 28; Indels 40; Gaps 7;

Qy 35 DLVHRRKPGGEFFYYTMMPKGTGYDDQRWILESGVDHYLKNKPSGRILNVYGTDFYF 94
||| : ||||| : | : | : | :
Db 572 DLSLPERKPG-----LQMGI--DCATWV---GVTG-----VVY----- 599

Qy 95 LTGAAVREMDHFETADGKTKYDISSKANGHPSRGK--NWGVWKDGEK 142

Db 600 -----ASRPDLFTFELDGEK--VKGVENGALFEGRSNGGSKGERR 641

RESULT 14
S00801
Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha-3 chain - human

C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 19-Apr-2002
C:Accession: S00801; S04019; A27397; S02275

R/Ovchinnikov, Y.A.; Monastyrskaya, G.S.; Broude, N.E.; Ushkaryov, Y.A.; Melkov, A.M.; Sverdlov, N.N.; Sverdlov, E.D.
 FEBS Lett. 233, 87-94, 1988
 A>Title: Family of human Na,K-ATPase genes. Structure of the gene for the catalytic subunit
 A'Reference number: S00801; MUID:88255304; PMID:2838329
 A'Accession: S00801
 A'Molecule type: DNA
 A'Residuals: 1-1013 <OVG>
 A'CROSS-references: EMBL:M37456
 R:Sverdlov, E.D.; Monastyrskaya, G.S.; Broude, N.E.; Ushkarev, Y.A.; Melkov, A.M.; Smirnov, N.N.; Ovchinnikov, Y.A.
 Dokl. Biochem. 297, 426-431, 1987
 A>Title: Family of human Na(+),K(+) -ATPase genes. Structure of the gene of isoform alpha
 A'Reference number: S04019
 A'Accession: S04019
 A'Molecule type: DNA
 A'Residuals: 1, 21H, 3-1013 <SVE1>
 A'CROSS-references: EMBL:X12910; NID:G28963
 A>Note: The authors translated the codon TTC for residue 283 as Ser and TCT for residue 284 as Phe. This paper is a translation of the Russian paper published in Dokl. Akad. Nauk SSSR, 1987, 283, 1013-1016.
 R:Sverdlov, E.D.; Monastyrskaya, G.S.; Broude, N.E.; Ushkaryov, Y.A.; Allikmets, R.L.; Matusik, M.B.; Sverdlov, V.E.; Modyanov, N.N.; Ovchinnikov, Y.A.
 FEBS Lett. 217, 275-278, 1987
 A>Title: The family of human Na,K-ATPase genes. No less than five genes and/or pseudogenes
 A'Reference number: A27397; MUID:87247232; PMID:3036582
 A'Accession: A27397
 A'Molecule type: mRNA
 A'Residuals: 243-434 <SVE2>
 A'CROSS-references: GB:M27570
 C/Genetics:
 A/Gene: GDB:ATPIA3
 A/CROSS-references: GDB:119713; OMIM:182350
 A/Map position: 19q13.2-19q13.2
 A/Introns: 2/3; 31/3; 51/3; 119/3; 157/3; 202/3; 242/1; 331/3; 338/1; 435/2; 479/3; 544/3
 C/Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain
 C/Keywords: ATP; heterodimer; hydrolase; ion transport; phosphoprotein; potassium transp
 F:86-110/Domain: transmembrane #status predicted <TM1>
 F:120-139/Domain: transmembrane #status predicted <TM2>
 F:140-280/Domain: intracellular #status predicted <INT2>
 F:281-303/Domain: transmembrane #status predicted <TM3>
 F:310-338/Domain: transmembrane #status predicted <TM4>
 F:339-776/Domain: intracellular #status predicted <INT3>
 F:777-773/Domain: ATPase nucleotide-binding domain homology <ATN>
 F:777-800/Domain: transmembrane #status predicted <TM5>
 F:839-864/Domain: transmembrane #status predicted <TM6>
 F:865-942/Domain: intracellular #status predicted <INT4>
 F:943-968/Domain: transmembrane #status predicted <TM7>
 F:969-1013/Domain: extracellular #status predicted <EXT>
 F:366/Active site: Asp (aspartylphosphate intermediate) #status predicted
 F:498/Binding site: ATP (Lys) #status predicted
 F:707,711,716/Active site: Asp, Asp, Lys #status predicted

Query Match 9.1%; Score 75; DB 1; Length 1013;
 Best Local Similarity 21.5%; Pred. No. 33;
 Matches 48; Conservative 23; Mismatches 56; Indels 96; Gaps 12;

QY 13 AIFGAASDN-PIADR-----KCIVISDGL-VNHER-KPQGEPPY-----YVY 52
 Db 421 AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSKVMRERKNKVAIPFNSNTKYL 480
 QY 53 MIPKGTEDDQDWILESVGGDH-----YKLNKFSGRYL----- 86
 Db 481 SIHETEDPNDNRVLLVMKGAPEIRILDRCSITLLOKQKPLDEEMKAFQNYLELGLGE 540
 QY 87 -VYGFDFYFL-----TAGAAVREMDHFKFTADG----- 113
 Db 541 RVLGFCHYLLPEEYPOQGFAP-DCDDVNFNTDNLFCVGLMSMIGPPRAAEPDVAVKCRSA 599
 QY 114 -----TGKYDISKANGHPSRGKNGWKCKEKFVTFVN 149
 Db 600 GIKVIMVGDHPITAKA-----IAKGVGIISGNE---TVED 633

RESULT 15

C24639

Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha-3 chain - rat
 N/Alternate names: Na+/K+-transporting ATPase alpha (III) chain
 C/Species: Rattus norvegicus (Norway rat)

C/Date: 30-Jun-1988 #sequence revision 23-Apr-1993 #text_change 19-Apr-2002
 C/Accession: C24639; S00514; E27180; A60470

R/Shull, G.B.; Greeb, J.; Lingrel, J.B.
 Biochemistry 25, 8125-8132, 1986

A>Title: Molecular cloning of three distinct forms of the Na+,K+-ATPase alpha-subunit f

A'Reference number: A90512; MUID:87128908; PMID:3028470

A'Accession: C24639

A'Molecule type: mRNA

A'Residuals: 1-1013 <SHU>

A/CROSS-references: EMBL:M14513; NID:G203030; PIDN:AAA40777.1; PID:G203031

A>Note: In the authors' translation 405-Ser is shown after residue 409 and, consequentl

R/Hara, Y.; Urayama, O.; Kawakami, K.; Nojima, H.; Nagamune, H.; Kojima, T.; Ohta, T.;

J. Biochem. 102, 43-58, 1987

A>Title: Primary structures of two types of alpha-subunit of rat brain Na(+),K(+)-ATPase

A'Reference number: S00460; MUID:88032933; PMID:2822682

A'Accession: S00514

A'Molecule type: mRNA

A'Residuals: 1-907,'C',909-1013 <HAR>

R/Herrera V.L.M.; Emanuel, J.R.; Ruiz-Opazo, N.; Levenson, R.; Nadai-Ginard, B.

J. Cell Biol. 105, 1855-1865, 1987

A>Title: Three differentially expressed Na,K-ATPase alpha subunit isoforms: structural

A'Reference number: A92749; MUID:88033255; PMID:2822726

A'Accession: B27180

A'Molecule type: mRNA

A'Residuals: 1, NL, 4-103,'R',105-113,'E',115-127,'G',129-148,'Q',150-151,'T',153-165,'D'

A/CROSS-references: EMBL:M28648; NID:G205633; PIDN:AAA41672.1; PID:G205634

A>Note: The authors translated the codon CAG for residue 149 as Glu, GGC for residue 19

R/Hsu, Y.M.; Guidotti, G.

Biochemistry 28, 569-573, 1989

A>Title: Rat brain has the alpha3 form of the (Na+,K+)ATPase.

A'Reference number: A60470; MUID:89229049; PMID:2540801

A'Accession: A60470

A'Molecule type: protein

A'Residuals: 117-132;586-595,'X',597-601 <HSU>

C/Comment: The alpha-3 form appears to be highly ouabain-inhibitable, as is alpha-2 but

C/Genetics:

A/Gene: NKAA3

C/Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain

C/Keywords: ATP; heterodimer; hydrolase; ion transport; phosphoprotein; potassium trans

F:86-110/Domain: transmembrane #status predicted <TM1>

F:120-139/Domain: transmembrane #status predicted <TM2>

F:140-280/Domain: intracellular #status predicted <INT2>

F:281-303/Domain: transmembrane #status predicted <TM3>

F:310-338/Domain: transmembrane #status predicted <TM4>

F:339-776/Domain: intracellular #status predicted <INT3>

F:777-773/Domain: ATPase nucleotide-binding domain homology <ATN>

F:777-800/Domain: transmembrane #status predicted <TM5>

F:839-864/Domain: transmembrane #status predicted <TM6>

F:865-942/Domain: intracellular #status predicted <INT4>

F:943-968/Domain: transmembrane #status predicted <TM7>

F:969-1013/Domain: extracellular #status predicted <EXT>

F:366/Active site: Asp (aspartylphosphate intermediate) #status predicted

F:498/Binding site: ATP (Lys) #status predicted

F:707,711,716/Active site: Asp, Asp, Lys #status predicted

Query Match 9.1%; Score 75; DB 2; Length 1013;
 Best Local Similarity 21.5%; Pred. No. 33;
 Matches 48; Conservative 23; Mismatches 56; Indels 96; Gaps 12;

QY 13 AIFGAASDN-PIADR-----KCIVISDGL-VNHER-KPQGEPPY-----YVY 52
 Db 421 AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSKVMRERKNKVAIPFNSNTKYL 480
 QY 53 MIPKGTEDDQDWILESVGGDH-----YKLNKFSGRYL----- 86
 Db 481 SIHETEDPNDNRVLLVMKGAPEIRILDRCSITLLOKQKPLDEEMKAFQNYLELGLGE 540

```
QY 87 -VYGFDFYFL-----TAGAAVREMDHPKFTADG----- 113
Db 541 RVLGFCHYILPEEQPKGFAP-DCDDVNFTTDNLCFVGLMSMIDPPRAAVPDVAVGKCRSA 599
QY 114 -----TKYDIDISSKANGHPRSRGKNWGMKDGKHYFTVEN 149
Db 600 GIKVIMVIGDHPITAKA-----IAKGVIISEGNE---TVED 633
```

Search completed: September 9, 2004, 15:29:21
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:15:29 ; Search time 23 Seconds
(without alignments)
344.116 Million cell updates/sec

Title: US-10-768-874-2
Perfect score: 825
Sequence: 1 MSITQSFVLTALFAGASD.....NWGVMDGKHFTVENCQE 152

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	99.5	12.1	637	SCAA_CHICK	Q92075 gallus gall
2	82	9.9	1597	GTF1_STRDO	P11001 streptococ
3	78	9.5	1129	YB95_YEAST	P38144 saccharomyc
4	77	9.3	173	CRBS_CYPAC	P10112 cyprinus ca
5	76.5	9.3	640	SCAA_RABIT	O97741 oryctolagus
6	75.5	9.2	437	GLGC_HAENI	P43796 haemophilus
7	75	9.1	1013	ALIA3_HUMAN	P13637 homo sapien
8	75	9.1	1013	ALIA3_RAT	P06687 rattus norv
9	73.5	8.9	613	PEPF_MYCPU	Q98QP0 mycoplasma
10	73	8.8	276	MYPR_CHICK	P23289 gallus gall
11	73	8.8	276	MYPR_PORCU	P47790 poephila gu
12	72.5	8.8	164	P825_ARATH	O27711 arabidopsis
13	72	8.7	616	ECAR_ECHCA	Q27711 arabidopsis
14	72	8.7	742	NSF_ARATH	Q90495 echis carin
15	71.5	8.7	358	SFR1_ARATH	Q90495 echis carin
16	71.5	8.7	358	RFL1_STAPF	O22315 arabidopsis
17	71.5	8.7	640	APRN_ENTHI	O8CNI7 staphylococ
18	71.5	8.7	656	ACSA_PHOCA	P20301 entameba h
19	71	8.6	362	OMPD_SALTY	O88040 rhodobacter
20	70.5	8.5	367	OMPC_ECOLI	P37592 salmonella
21	70.5	8.5	473	6PGD_BUCAP	P06996 escherichia
22	70.5	8.5	1139	INAI_CAEEL	Q92H39 buchnera ap
23	70	8.5	330	MTN3_HAFAE	Q03600 caenorhabdi
24	70	8.5	441	XYLA_RHILO	P20589 haemophilus
25	69.5	8.4	275	LM22_VIBPA	Q98CR8 rhizobium l
26	69.5	8.4	603	GLMS_THETH	Q56714 vibrio para
27	69.5	8.4	778	RHP9_SCHPO	Q56213 t glucosami
28	68.5	8.3	185	COMI_DICDI	P87074 schizosacch
29	68.5	8.3	698	SCAA_RAT	Q03380 dictyosteli
30	68.5	8.3	711	TBBI_NEIMB	P37089 rattus norv
31	68.5	8.3	910	SYL_NEIMA	Q09057 neisseria m
32	68	8.2	480	DISA_TRIGA	Q9JW39 neisseria m
33	68	8.2	597	GLMS_PYRAB	P15503 trimeresuri
					Q9V249 p glucosami

34 68 8.2 1507 1 CADE_DROME
35 67.5 8.2 340 1 LPSE_RHIME
36 67.5 8.2 365 1 FENR_MESCR
37 67.5 8.2 398 1 TRMU_BRUME
38 67.5 8.2 398 1 TRMU_BRUSU
39 67.5 8.2 445 1 6PGD_CITAM
40 67.5 8.2 445 1 6PGD_CITFR
41 67.5 8.2 485 1 PDII_CAEEL
42 67.5 8.2 505 1 MOT6_HUMAN
43 67.5 8.2 565 1 SYG_METTH
44 67.5 8.2 876 1 SYL_NEIMB
45 67.5 8.2 2333 1 POLG_FMDVI

ALIGNMENTS

RESULT 1
SCAA_CHICK
ID SCAA_CHICK STANDARD; PRT; 637 AA.
AC Q92075; P70095; Q98941;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Amiloride-sensitive sodium channel alpha-subunit (Epithelial Na+ channel alpha subunit) (Alpha ENAC) (Nonvoltage-gated sodium channel 1 alpha subunit) (SCN5A) (Alpha NaCh).
DE channel alpha subunit (SCN5A) (Alpha NaCh).
GN SCNN1A OR ENAC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=Isa brown; TISSUE=Cochlea;
EX MEDLINE=97157073; PubMed=9003454;
RA Killick R., Richardson G.;
RT "Isolation of chicken alpha ENAC splice variants from a cochlear cDNA library."
RL Biochim. Biophys. Acta 1350:33-37(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Goldstein O., Asher C., Garty H.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Sodium permeable non-voltage-sensitive ion channel inhibited by the diuretic amiloride. Mediates the electrodiffusion of the luminal sodium (and water, which follows osmotically) through the apical membrane of epithelial cells. Controls the reabsorption of sodium in kidney, colon, lung and sweat glands.
CC -!- SUBUNIT: Heterotetramer of two alpha, one beta and one gamma subunit. A delta subunit can replace the alpha subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q92075-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q92075-2; Sequence=VSP_006196, VSP_006197;
CC TISSUE SPECIFICITY: The long isoform has been found in cochlea, colon, and cartilage. The short isoform is only found in cochlea.
CC -!- PTM: Ubiquitinated; this targets individual subunits for proteasome-mediated degradation (By similarity).
CC -!- SIMILARITY: Belongs to the amiloride-sensitive sodium channel family.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-15 is the initiator.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its


```

SEQUENCE FROM N.A.
TISSUE=Lens;
MEDLINE=88000706; PubMed=3307924;
Chang T., Chang W.-C.;
"Cloning and sequencing of a carp beta s-crystallin cDNA.";
Biochim. Biophys. Acta 910:89-92(1987).
CC !- FUNCTION: Crystallins are the dominant structural components of
the vertebrate eye lens.
CC !- DOMAIN: Has a two-domain beta-structure, folded into four very
similar greek key motifs.
CC !- SIMILARITY: Belongs to the beta/gamma-crystallin family.
CC !- SIMILARITY: Contains 4 beta/gamma-crystallin 'Greek key' domains.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M26834; AAA49206.1; -.
FIR; S07146; S07146.
HSSP; P06504; IAVH.
InterPro; IPR001064; Crystallin.
Pfam; PF00030; crystall; 2.
PRINTS; PR01367; BGCYCRYSTALLIN.
SMART; SM00247; XTALbg; 2.
PROSITE; PS50915; CRYSTALLIN_BETAGAMMA; 4.
Eye lens protein; Repeat.
INIT MET 0 0
FT DOMAIN 1 39 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 1.
FT DOMAIN 40 82 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 2.
FT DOMAIN 83 88 CONNECTING PEPTIDE.
FT DOMAIN 89 129 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 3.
FT DOMAIN 130 172 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 4.
FT SEQUENCE 173 AA; 20697 MW; 2E5DBA95453D2897 CRC64;
-----
Query Match 9.3%; Score 77; DB 1; Length 173;
Best Local Similarity 26.7%; Pred. No. 1.3;
Matches 36; Conservative 18; Mismatches 41; Indels 40; Gaps 9;
QY 29 IVISGDLVHHRKPGQHPFYVYVIMPKGTVEYDQRW-----ILEYGGDPHYKL 77
DB 35 IRVESGAWYIYER---PNPMGQYVLTGRGEPDYQRFWMGLNDRLCCKMIFHVSQSEYKI 91
QY 78 ----KNKSGRYLVY-----GTFDYFLTAGAAVREMDHEKFTADGTGKYDISSKANGH 126
DB 92 QLYDKGDFGQ--VVESTEDCESVDVDRFT-----REVHSCK-VLDGIWIF-----YEH 137
QY 127 PRRSGKNMGVWKDGE 141
DB 138 PNYRGRQY-LLEKGE 151
-----
RESULT 5
SCAA RABIT
ID SCAA RABIT STANDARD; PRT; 640 AA.
AC O97741; Q9N133;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-WAR-2004 (Rel. 43, Last annotation update)
DE Amiloride-sensitive sodium channel alpha-subunit (Epithelial Na+
channel alpha subunit) (Alpha ENaC) (Nonvoltage-gated sodium channel 1
alpha subunit) (SCN9A) (Alpha NaCh).
DE SCNN1A.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CC NCBI_taxid=9986;
RN [1]
RR H. Wiener H. Pluss H.:
RP Kudlacek O., Weisz E.,

```

"The rabbit epithelial sodium channel."
Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 81-435 FROM N.A.
RA Velazquez H., Silva T.C., Andujar E., Jaffer A., Ortiz D.;
RA "The rabbit DCT does not express amiloride sensitive sodium channel."
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Sodium permeable non-voltage-sensitive ion channel
CC inhibited by the diuretic amiloride. Mediates the electrodifusion
CC of the luminal sodium (and water, which follows osmotically)
CC through the apical membrane of epithelial cells. Controls the
CC reabsorption of sodium in kidney, colon, lung and sweat glands.
CC Also plays a role in taste perception.
CC -!- SUBUNIT: Heterotrimer of two alpha, one beta and one gamma
CC subunit. A delta subunit can replace the alpha subunit. Interacts
CC with the WW domains of NEDD4, WWP1 and WWP2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: Ubiquitinated; this targets individual subunits for
CC proteasome-mediated degradation (By similarity).
CC -!- SIMILARITY: Belongs to the amiloride-sensitive sodium channel
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ132108; CAA10571.1; -;
CC EMBL; AF229025; AAF43680.1; -;
CC InterPro; IPR004724; Enac.
CC Pfam; PF00858; ASC; 1.
CC PRINTS; PR01078; AMINACHANNEL.
CC TIGRFS; TIGR00859; ENAC; 1.
CC PROSITE; PS01206; ASC; 1.
CC Ion transport; Sodium transport; Ionic channel; Transmembrane;
CC Glycoprotein; Ubl conjugation; Sodium channel.
CC DOMAIN 1 65 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 66 86 POTENTIAL.
CC FT DOMAIN 87 539 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 540 560 POTENTIAL.
CC FT TRANSMEM 561 640 POTENTIAL.
CC FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 222 222 S -> P (IN REF. 2).
CC FT CONFLICT 264 265 AF -> TS (IN REF. 2).
CC FT CONFLICT 372 373 TQ -> AE (IN REF. 2).
CC SEQUENCE 640 AA; 72734 MW; 17C84C100F69B133 CRC64;
Query Match 9.3%; Score 76.5; DB 1; Length 640;
Best Local Similarity 30.5%; Pred. No. 6.2;
Matches 29; Conservative 10; Mismatches 31; Indels 25; Gaps 5;
QY 49 YVYVMPKGYDDQR---WILESGDHYKLNKFSGRVLYGTF-----DYF 94
Db 410 YIYFPEGEYCYDKHNSW-----GYCYKLODAFSSDL--GCTKCRKCSVTNYE 462
QY 95 LTAG----AAVREMDHKFTADGTGKIDISSKANG 125
Db 463 LSAGYSRWPSVTSQDWVFMQLSQNNYVTSNKENG 497
RESULT 6
GLGC_HAEIN STANDARD; PRT; 437 AA.
ID GLGC_HAEIN

P43796;
01-NOV-1995 (Rel. 32, Created)
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc PPase).
GN GLGC OR H11359.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=93550630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd".
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -!- PATHWAY: Glycogen biosynthesis; first step.
CC -!- SIMILARITY: Belongs to the bacterial/plant glucose-1-phosphate
CC adenylyltransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32815; AAC23006.1; ALT_INIT.
CC PIR; B64119; B64119.
CC TIGR; H11359; -;
CC HAMAP; MF_00624; -; 1.
CC InterPro; IPR005836; ADP_Glu_Pyrop.
CC InterPro; IPR005835; NTP_transferase.
CC Pfam; PF00483; NTP_transferase; 1.
CC PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
CC PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
CC PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
CC Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
CC Complete proteome.
CC SEQUENCE 437 AA; 49107 MW; E5E2284ACE5AD1D9 CRC64;
Query Match 9.2%; Score 75.5; DB 1; Length 437;
Best Local Similarity 26.7%; Pred. No. 5.1;
Matches 27; Conservative 16; Mismatches 47; Indels 11; Gaps 3;
QY 45 QRFYYVYVMPKGYDDQR---WILESGDHYKLNKFSGRVLYGTFDYFLTAGAAV 101
Db 92 QERGEVDMFAPARQQIDSDTWYRGTDADAVYQNNAIKKNHYPKYLILAGDHIYKODYSV 151
QY 102 REMDHKFTADGTGKIDISSKANGHPRGRKNGKNGVMDKE 141
Db 152 MLMDHVNSGAKCTVGCIEV-----PRSEAEHFGVWAYNE 185
RESULT 7
ALIA3_HUMAN
ID ALIA3_HUMAN STANDARD; PRT; 1013 AA.
AC P13637; Q16732; Q16735; Q969K5;
DT 01-JAN-1990 (Rel. 13, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Sodium/potassium-transferring ATPase alpha-3 chain (EC 3.6.3.9)
 GN (Sodium pump 3) (Na+/K+ ATPase 3) (Alpha (III)).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86255304; PubMed=2838329;
 RA Ovchinnikov Y.A., Monastyrskaya G.S., Broude N.E., Ushkaryov Y.A.,
 RA Melkov A.M., Smirnov Y.V., Malyshev I.V., Allikmets R.L.,
 RA Kostina M.B., Dulubova I.E., Kiyatkin N.I., Grishin A.V.,
 RA Modyanov N.N., Ovchinnikov Y.A.,
 RA Modyanov N.N., Sverdlov E.D.;
 RT "Family of human Na+, K+-ATPase genes. Structure of the gene for the
 RT catalytic subunit (alpha III-form) and its relationship with
 RT structural features of the protein.";
 RL FEBS Lett. 233:87-94 (1988).
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Sverdlov E.D., Monastyrskaya G.S., Broude N.E., Ushkarev Y.A.,
 RA Melkov A.M., Smirnov Y.V., Malyshev I.V., Allikmets R.L.,
 RA Kostina M.B., Dulubova I.E., Kiyatkin N.I., Grishin A.V.,
 RA Modyanov N.N., Ovchinnikov Y.A.,
 RT "Family of human Na+(K+)-ATPase genes. Structure of the gene of
 RT isoform alpha-III.";
 RL Dokl. Akad. Nauk SSSR 297:1488-1494 (1987).
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.B.,
 RA Rana S.S., Loquellano N.B., Toshiyuki S., Carninci P., Prange C.,
 RA Bosak S.A., McWane P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
 RN
 RP SEQUENCE OF 120-387; 494-538 AND 545-1013 FROM N.A.
 RX MEDLINE=87162481; PubMed=3030810;
 RA Ovchinnikov Y.A., Monastyrskaya G.S., Broude N.E., Allikmets R.L.,
 RA Ushkaryov Y.A., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
 RA Dulubova I.E., Petrukhin K.E., Gryshin A.V., Sverdlov V.E.,
 RA Kiyatkin N.I., Kostina M.B., Modyanov N.N., Sverdlov E.D.;
 RT "The family of human Na+, K+-ATPase genes. A partial nucleotide
 RT sequence related to the alpha-subunit.";
 RL FEBS Lett. 213:73-80 (1987).
 RN
 RP ERATUM.
 RA Ovchinnikov Y.A., Monastyrskaya G.S., Broude N.E., Allikmets R.L.,
 RA Ushkaryov Y.A., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
 RA Dulubova I.E., Petrukhin K.E., Gryshin A.V., Sverdlov V.E.,
 RA Kiyatkin N.I., Kostina M.B., Modyanov N.N., Sverdlov E.D.;
 RL FEBS Lett. 214:375-375 (1987).
 RN
 RP SEQUENCE OF 243-434 FROM N.A.

RX MEDLINE=87247232; PubMed=3036582;
 RA Sverdlov E.D., Monastyrskaya G.S., Broude N.E., Ushkaryov Y.A.,
 RA Allikmets R.L., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
 RA Dulubova I.E., Petrukhin K.E., Grishin A.V., Kiyatkin N.I.,
 RA Kostina M.B., Sverdlov V.E., Modyanov N.N., Ovchinnikov Y.A.;
 RT "The family of human Na+, K+-ATPase genes. No less than five genes
 RT and/or pseudogenes related to the alpha-subunit.";
 RL FEBS Lett. 217:275-278 (1987).
 CC
 CC -!- FUNCTION: This is the catalytic component of the active enzyme,
 CC which catalyzes the hydrolysis of ATP coupled with the exchange of
 CC sodium and potassium ions across the plasma membrane. This action
 CC creates the electrochemical gradient of sodium and potassium ions,
 CC providing the energy for active transport of various nutrients.
 CC -!- CATALYTIC ACTIVITY: ATP + H2O + Na+(In) + K+(Out) = ADP +
 CC phosphate + Na+(Out) + K+(In).
 CC -!- SUBUNIT: Composed of three subunits: alpha (catalytic), beta and
 CC gamma.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IIC.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M37457; AAA51798.1; JOINED.
 DR EMBL; M37436; AAA51798.1; JOINED.
 DR EMBL; M37437; AAA51798.1; JOINED.
 DR EMBL; M37438; AAA51798.1; JOINED.
 DR EMBL; M37462; AAA51798.1; JOINED.
 DR EMBL; M37439; AAA51798.1; JOINED.
 DR EMBL; M37440; AAA51798.1; JOINED.
 DR EMBL; M37441; AAA51798.1; JOINED.
 DR EMBL; M37442; AAA51798.1; JOINED.
 DR EMBL; M37443; AAA51798.1; JOINED.
 DR EMBL; M37444; AAA51798.1; JOINED.
 DR EMBL; M37445; AAA51798.1; JOINED.
 DR EMBL; M37447; AAA51798.1; JOINED.
 DR EMBL; M37448; AAA51798.1; JOINED.
 DR EMBL; M37449; AAA51798.1; JOINED.
 DR EMBL; M37450; AAA51798.1; JOINED.
 DR EMBL; M37451; AAA51798.1; JOINED.
 DR EMBL; M37452; AAA51798.1; JOINED.
 DR EMBL; M37453; AAA51798.1; JOINED.
 DR EMBL; M37454; AAA51798.1; JOINED.
 DR EMBL; M37455; AAA51798.1; JOINED.
 DR EMBL; M37456; AAA51798.1; JOINED.
 DR EMBL; X12910; CAA31390.1; JOINED.
 DR EMBL; X12911; CAA31390.1; JOINED.
 DR EMBL; X12912; CAA31390.1; JOINED.
 DR EMBL; X12913; CAA31390.1; JOINED.
 DR EMBL; X12914; CAA31390.1; JOINED.
 DR EMBL; X12915; CAA31390.1; JOINED.
 DR EMBL; X12916; CAA31390.1; JOINED.
 DR EMBL; X12917; CAA31390.1; JOINED.
 DR EMBL; X12919; CAA31390.1; JOINED.
 DR EMBL; X12920; CAA31390.1; JOINED.
 DR EMBL; X12921; CAA31390.1; JOINED.
 DR EMBL; X12922; CAA31390.1; JOINED.
 DR EMBL; X12923; CAA31390.1; JOINED.
 DR EMBL; M28286; AAA52285.1; JOINED.
 DR EMBL; M28284; AAA52285.1; JOINED.
 DR EMBL; M28285; AAA52285.1; JOINED.
 DR EMBL; M28293; AAA52286.1; JOINED.
 DR EMBL; M28287; AAA52286.1; JOINED.
 DR EMBL; M35821; AAA52286.1; JOINED.
 DR EMBL; M35822; AAA52286.1; JOINED.
 DR EMBL; M28289; AAA52286.1; JOINED.
 DR EMBL; M28290; AAA52286.1; JOINED.

```
DR EMBL; M28291; AAA52286.1; JOINED.
DR EMBL; M28292; AAA52286.1; JOINED.
DR EMBL; M27577; AAA58380.1; -.
DR EMBL; M27570; AAA58380.1; JOINED.
DR EMBL; M27573; AAA58380.1; JOINED.
DR EMBL; BC009282; AAH09282.1; -.
DR EMBL; BC009394; AAH09394.1; -.
DR EMBL; BC015566; AAH15566.1; -.
DR PIR; S00801; S00801.
DR HSSP; P04191; 1EUL.
DR Genew; HGNC:801; ATP1A3.
DR MIM; 192350; -.
DR GO; GO:0005890; C:sodium/potassium-exchanging ATPase complex; TAS.
DR GO; GO:0005391; F:sodium/potassium-exchanging ATPase activity; TAS.
DR GO; GO:0006810; P:transport; TAS.
DR InterPro; IPR001757; ATPase E1-E2.
DR InterPro; IPR006069; Cation ATPase.
DR InterPro; IPR006068; Cation ATPase C.
DR InterPro; IPR004014; Cation ATPase N.
DR InterPro; IPR008250; E1-E2 ATPase_Reg.
DR InterPro; IPR005834; Hydrolyase.
DR InterPro; IPR005775; Na/K ATPase_alph.
DR Pfam; PF00689; Cation ATPase C; 1.
DR Pfam; PF00690; Cation ATPase N; 1.
DR Pfam; PF00122; E1-E2 ATPase; 1.
DR Pfam; PF00702; Hydrolyase; 1.
DR PRINTS; PR00119; CATATPASE.
DR PRINTS; PR00121; NAKATPASE.
DR TIGRFS; TIGR01106; ATPase-IIC X-K; 1.
DR TIGRFS; TIGR01494; ATPase P-type; 5.
DR PROSITE; PS00154; ATPase E1_E2; 1.
DR PROSITE; PS00154; Sodium/potassium transport; Transmembrane; Phosphorylation;
KW Magnesium; Metal-binding; ATP-binding; Multigene family.
FT DOMAIN 1 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 98 POTENTIAL.
FT DOMAIN 99 121 LUMENAL (POTENTIAL).
FT TRANSMEM 122 142 POTENTIAL.
FT DOMAIN 143 278 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 279 298 POTENTIAL.
FT DOMAIN 299 310 LUMENAL (POTENTIAL).
FT TRANSMEM 311 328 POTENTIAL.
FT DOMAIN 329 762 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 763 782 POTENTIAL.
FT DOMAIN 783 792 LUMENAL (POTENTIAL).
FT TRANSMEM 793 813 POTENTIAL.
FT DOMAIN 814 833 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 834 856 POTENTIAL.
FT DOMAIN 857 908 LUMENAL (POTENTIAL).
FT TRANSMEM 909 928 POTENTIAL.
FT DOMAIN 929 941 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 942 960 POTENTIAL.

Query Match 9.1%; Score 75; DB 1; Length 1013;
Best Local Similarity 21.5%; Pred. No. 14;
Matches 48; Conservative 23; Mismatches 56; Indels 96; Gaps 12;

QY 13 ALFGRASDN-PIADR-----KCVILSDGDL-VNHER-KPGQRPFY-----YVY 52
DB 421 AVFKGGQDNPVLKRDVADGASESALKKCIELSSGVSVKLMREKKVAEIPFNSTKYQL 480
QY 53 MIPKGPTEYDDQWILESVGGDH-----KCVILSDGDL-VNHER-KPGQRPFY-----YVY 86
DB 481 SIHETEDPNDNRVLLVMKGAPEILDRCSITLLQKGEQPLDEBMEKAFQNAVYELGLGGE 540
QY 87 -YVGTEDYFL-----TAGAAVEMDFKTTADG-----YVGTEDYFL-----YVY 113
DB 541 RVLGFCYHYLPPEQFPKGFAP-DCDDVNFPTDMLCFVGLMSMTDPPRAAVPDVAVGKCRSA 599
QY 114 -----TGKYDISSKANGPRGRGNKGNWGMKDKGKHFTVEN 149
DB 600 GIKVIMVTGDPITAKA-----IAKGVGLIISGNE---TVED 633

RESULT 8
ALA3 RAT STANDARD; PRT; 1013 AA.
ID P06687; Q61732; Q921G6;
DT 01-JAN-1988 (Rel. 06, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sodium/potassium-transporting ATPase alpha-3 chain (EC 3.6.3.9)
DE (Sodium pump 3) (Na+/K+ ATPase 3) (Alpha(III)).
GN ATP1A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87128908; PubMed=3028470;
RA Shull G.E., Greeb J., Lingrel J.B.;
RT "Molecular cloning of three distinct forms of the Na+,K+-ATPase
RT alpha-subunit from rat brain.";
RL Biochemistry 25:8125-8132(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=88032933; PubMed=2822682;
RA Hara Y., Urayama O., Kawakami K., Nojima H., Nagamune H., Kojima T.,
RA Ohta T., Nagano K., Nakao M.;
RT "Primary structures of two types of alpha-subunit of rat brain
RT Na+,K+-ATPase deduced from cDNA sequences.";
RL J. Biochem. 102:43-58(1987).
RN [3]
RP SEQUENCE OF 1-930 FROM N.A.
RX TISSUE=Brain, and Liver;
RX MEDLINE=88033255; PubMed=2822726;
RA Herrera V.L.M., Emanuel J.R., Ruiz-Opazo N., Levenson R.,
RA Nadal-Ginard B.;
RT "Three differentially expressed Na,K-ATPase alpha subunit isoforms:
RT structural and functional implications.";
RL J. Cell Biol. 105:1855-1865(1987).
RN [4]
RP SEQUENCE OF 243-434 FROM N.A.
RX MEDLINE=87247232; PubMed=3036582;
RA Sverdlov E.D., Monastyrskaya G.S., Broude N.E., Ushkaryov Y.A.,
RA Allikmets R.L., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
RA Dulobova I.E., Petrukhin K.E., Grishin A.V., Kijatkin N.I.,
RA Kosina M.B., Sverdlov V.E., Modyanov N.N., Ovchinnikov Y.A.;
RT "The family of human Na+,K+-ATPase genes. No less than five genes
RT and/or pseudogenes related to the alpha-subunit.";
RL FEBS Lett. 217:275-278(1987).
CC -!- FUNCTION: This is the catalytic component of the active enzyme,
CC which catalyzes the hydrolysis of ATP coupled with the exchange of
CC sodium and potassium ions across the plasma membrane. This action
CC creates the electrochemical gradient of sodium and potassium ions,
CC providing the energy for active transport of various nutrients.
CC -!- CATALYTIC ACTIVITY: ATP + H(2O) + Na(+) (in) + K(+) (out) = ADP +
CC phosphate + Na(+) (out) + K(+) (in).
CC -!- SUBUNIT: Composed of three subunits: alpha (catalytic), beta and
CC gamma.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14513; AAA40777.1; -.
CC EMBL; X05883; CAA29307.1; -.
CC EMBL; M28648; AAA41672.1; -.
CC
```


36 HEALTHTEQLIETFF---SKNYQDYFYLLDVI-----HAFQVYVYGTASFFFLYG 82

Db 36 HEALTHTEQLIETFF---SKNYQDYFYLLDVI-----HAFQVYVYGTASFFFLYG 82
Qy 99 A-----AVREM-DHEKFTADGTGKYDICKKANGHPRSRG 131
Db 83 ALLLAGPYTTCAVRQIFDRTTCGKG---LSATVTGGPKGRG 124

RESULT 11

MYPR_POEGU
ID MYPR_POEGU STANDARD; PRT; 276 AA.
AC P47790;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myelin proteolipid protein (PLP) (Lipophilin).
GN PLP1 OR PLP.
OS *Popilia guttata* (Zebra finch) (*Taeniopygia guttata*).
OC *Eukaryota*; *Metazoa*; *Chordata*; *Cranialata*; *Vertebrata*; *Euteleostomi*;
OC *Archosauria*; *Aves*; *Neognathae*; *Passeriformes*; *Estrildidae*;
OC *Estrildinae*; *Taeniopygia*.
OX NCBI_TaxID=59729;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95098210; PubMed=7528351;
RA Campagnoni C.W., Kamp K., Mason B., Handley V.W., Campagnoni A.T.;
RT "Isolation and characterization of a cDNA encoding the zebra finch
myelin proteolipid protein."
RL Neurochem. Res. 19:1061-1065(1994).
CC -!- FUNCTION: THIS IS THE MAJOR MYELIN PROTEIN FROM THE CENTRAL
NERVOUS SYSTEM. IT PLAYS AN IMPORTANT ROLE IN THE FORMATION OR
MAINTENANCE OF THE MULTILAMELLAR STRUCTURE OF MYELIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the myelin proteolipid protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X51661; CAA33839.1; --
CC PIR; S17600; S17600.
CC InterPro; IPR001614; Myelin PLP.
CC Pfam; PF01275; Myelin PLP; 1.
CC PRINTS; PR00214; MYELINPLP.
CC SMART; SM00002; PLP; 1.
CC PROSITE; PS00575; MYELIN PLP 1; 1.
CC PROSITE; PS01004; MYELIN PLP 2; 1.
CC Myelin; Transmembrane; Structural protein; Lipoprotein; Palmitate.
CC INIT MET 0 0 BY SIMILARITY.
CC DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 10 35 I (POTENTIAL).
CC DOMAIN 36 58 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 59 87 II (POTENTIAL).
CC DOMAIN 88 150 III (POTENTIAL).
CC TRANSMEM 151 177 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 238 267 IV (POTENTIAL).
CC TRANSMEM 268 276 CYTOPLASMIC (POTENTIAL).
CC LIPID 5 5 S-palmitoyl cysteine (By similarity).
CC LIPID 6 6 S-palmitoyl cysteine (By similarity).
CC LIPID 9 9 S-palmitoyl cysteine (By similarity).
CC LIPID 108 108 S-palmitoyl cysteine (By similarity).
CC LIPID 138 138 S-palmitoyl cysteine (By similarity).
CC LIPID 140 140 S-palmitoyl cysteine (By similarity).
CC LIPID 183 227 S-palmitoyl cysteine (By similarity).
CC DISULFID 200 219 BY SIMILARITY.
CC SEQUENCE 276 AA; 29992 MW; 2FC540D5B649663F CRC64;
Query Match 8.8%; Score 73; DB 1; Length 276;
Best Local Similarity 23.8%; Pred. No. 5.4;
Matches 25; Conservative 17; Mismatches 35; Indels 28; Gaps 5;

RESULT 10

MYPR_CHICK
ID MYPR_CHICK STANDARD; PRT; 276 AA.
AC P23289;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myelin proteolipid protein (PLP) (Lipophilin).
GN PLP1 OR PLP.
OS *Gallus gallus* (Chicken).
OC *Eukaryota*; *Metazoa*; *Chordata*; *Cranialata*; *Vertebrata*; *Euteleostomi*;
OC *Archosauria*; *Aves*; *Neognathae*; *Galliformes*; *Phasianinae*;
OC *Gallus*.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Brain;
RX MEDLINE=92126243; PubMed=1722981;
RA Schless F., Scofield W.;
RT "Evolution of the myelin integral membrane proteins of the central
nervous system."
RL Biol. Chem. Hoppe-Seyler 372:865-874(1991).
CC -!- FUNCTION: THIS IS THE MAJOR MYELIN PROTEIN FROM THE CENTRAL
NERVOUS SYSTEM. IT PLAYS AN IMPORTANT ROLE IN THE FORMATION OR
MAINTENANCE OF THE MULTILAMELLAR STRUCTURE OF MYELIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the myelin proteolipid protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X51661; CAA33839.1; --
CC PIR; S17600; S17600.
CC InterPro; IPR001614; Myelin PLP.
CC Pfam; PF01275; Myelin PLP; 1.
CC PRINTS; PR00214; MYELINPLP.
CC SMART; SM00002; PLP; 1.
CC PROSITE; PS00575; MYELIN PLP 1; 1.
CC PROSITE; PS01004; MYELIN PLP 2; 1.
CC Myelin; Transmembrane; Structural protein; Lipoprotein; Palmitate.
CC INIT MET 0 0 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 1 9 I (POTENTIAL).
CC TRANSMEM 10 35 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 36 58 II (POTENTIAL).
CC TRANSMEM 59 87 III (POTENTIAL).
CC DOMAIN 88 150 IV (POTENTIAL).
CC TRANSMEM 151 177 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 238 267 S-palmitoyl cysteine (By similarity).
CC TRANSMEM 268 276 S-palmitoyl cysteine (By similarity).
CC LIPID 5 5 S-palmitoyl cysteine (By similarity).
CC LIPID 6 6 S-palmitoyl cysteine (By similarity).
CC LIPID 9 9 S-palmitoyl cysteine (By similarity).
CC LIPID 108 108 S-palmitoyl cysteine (By similarity).
CC LIPID 138 138 S-palmitoyl cysteine (By similarity).
CC LIPID 140 140 S-palmitoyl cysteine (By similarity).
CC LIPID 183 227 S-palmitoyl cysteine (By similarity).
CC DISULFID 200 219 BY SIMILARITY.
CC SEQUENCE 276 AA; 29992 MW; 2FC540D5B649663F CRC64;
Query Match 8.8%; Score 73; DB 1; Length 276;
Best Local Similarity 23.8%; Pred. No. 5.4;
Matches 25; Conservative 17; Mismatches 35; Indels 28; Gaps 5;

Qy 39 HERKFGQFFPYVYVIMPKGTEYDDQDWILSVGDHVKLNKPSGRVLVGTGDFYTAG 98

Query Match 8.8%; Score 73; DB 1; Length 276;
Best Local Similarity 24.8%; Pred. No. 5.4;

Matches	26;	Conservative	15;	Mismatches	36;	Indels	28;	Gaps	5;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 39 HERKPGQEPPYYVMIPIKGTEYDDQRWILESGVDHYKLNKPSGRYLVTGYDYFLTAG 98
||| ||| | | | | | | | | | | |
Dd 36 HEALTGTQLIETVF---SKNYQDVEFLIDVIHG-----FOYRIVGTAFFFLVG 82

Qy	99	A	-----	AVREM-DHFKFTADGTGKYDISSKANGHPSRG	131
			:	:	:
			:	:	:
Db	83	ALLLAEGFYTTGAVRQIFGDRYTTICGKG---		LSATVTGPKKRG	124

RESULT 12
F825_ARATH
ID_F825 ARATH STANDARD; PRT; 164 AA.

DT	15-DEC-1998	(Rel. 37, Created)
DT	15-DEC-1998	(Rel. 37, Last sequence update)
DT	28-FEB-2003	(Rel. 41, Last annotation update)

GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Tracheophytes
OC ATIG60740 OR F8A5.25.
DB ATIG60740 OR F8A5.25.
PC Biophysical Properties

OC eumecophyta; magnoliophyta; eudicotyledons; core eudicotyledons;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCEI TaxID=3702;
RN [1] _ RN

SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul

Bushler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii

B.A., C.E. Gosselink A.D., Haas B., Hansen N.F., Hughes B.
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khayk
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li

Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick
 W.C., Wu C.H., Wu Z.H., Wuos O.S., Mallik K., Marziani

Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Wais
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
"Sequence and analysis of chromosome 1 of the plant *Arabidopsis*

Nature 408:816-820(2000).
 CC -1- SIMILARITY: Belongs to the peroxiredoxin 2 family.
 CC -----
 CC -----
 CC -----

between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on the use of the data produced through this collaboration by non-profit institutions as long as its content

CC entities requires a license agreement (See <http://www.isb-sib.ch>),
CC or send an email to license@isb-sib.ch.
CC
CC
CC

DR PIR; G96632; G96632.
DR HSSP; P30044; 1H4O.
DR InterPro; IPR000866; AhpC-TSA.

KW Hypothetical protein.
 SQ SEQUENCE 164 AA; 17841 MW; 8D934C52C7311E84 CRC64;

Best Local Similarity 27.8%; Pred. No. 3.4;
Matches 25; Conservative 14; Mismatches 36; Indels

63 GRAEELKSGIDEILICFVSNDPFVMKAWGKTYQENKHVKFVADGSGEYTH

Qy 122 KANGHPRSRGKNGVMKDGEKHFTVENCQ 151
 Db 123 KGLG---IRSRFALLLDNLK--VTVANVE 147

RESULT 13
ECAR_ECHCA
ID_ECAR ECHCA STANDARD: PRT: 616 AA

AC	Q90495;	28-FEB-2003	(Rel. 41, Created)
DT		28-FEB-2003	(Rel. 41, Last sequence update)
DT		10-OCT-2003	(Rel. 42, Last annotation update)

DE
OS
OC
OC
OC

Ecarin precursor (EC 3.4.24.-).
Echis carinatus (Saw-scaled viper).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata.
Lepidosauria; Squamata; Sclerodlossa; Serpentes.
Lepidosauria; Squamata; Sclerodlossa; Serpentes.

OC
OX
RN
RP

viperidae; viperinae; Echis.
NCBI_TaxID=40353;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 191-228

RC TISSUE=Venom gland;
 RX MEDLINE=55151760; PubMed=7849037;
 RA Nishida S., Fujita T., Kohno N., Atoda H.:
 393-413; 448-453; 459-474; 553-561; 574-591

RT "CDNA cloning and deduced amino acid sequence
RT (ecarin) from Kenyan *Echis carinatus* venom."
RL Biochemistry 34:1771-1778(1995).

RP CHARACTERIZATION.
RC TISSUE=Venom;
RX MEDLINE=76260130; PubMed=956136;

RT The mechanism of activation of bovine proth-

-I- FUNCTION: Catalyzes the conversion of prothrombin through formation of a thrombin metalloproteinase activity.

-! COORDINATOR: BINDS 1 ZINC ION AND 1 CALCIUM ION (SIMILARITY).
CC
CC
CC
CC
-! SUBUNIT: Monomer.
CC
-! SIMILARITY: Belongs to peptidase family

 . CONTACT: CONTACTS & DISSEMINATION

 This SWISS-PROT entry is copyright. It is provided
 between the Swiss Institute of Bioinformatics
 CC
 CC
 CC
 CC
 CC

use by non-profit institutions as long as the statement is not removed. The statement requires a license agreement (See here)

CC
DR EMBL; D32212; BAA06910.1; -.
DR PIR; A55796; A55796

DR MEROFS; M12.151; -;
DR InterPro; IPR006586; ADAM cysteine.
DR InterPro; IPR001762; Disintegrin.

InterPro; IPR000287; Peptidase_M12B_N.
Pfam; PF00200; disintegrin; 1.

DR PFam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.

	DR	SMART; SM00050; DISIN; 1.
	DR	PROSITE; PS50215; ADAM MEPRO; 1.
	DR	PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.


```

DR HSP; P18708; INSE.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003599; AAA ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR009010; ASP_decarb_fold.
DR InterPro; IPR003338; ATPaseVAT_N.
DR InterPro; IPR004201; Cdc48_2.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02933; Cdc48_2; 1.
DR Pfam; PF02359; Cdc48_N; 1.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00674; AAA; 1.
KW Transport; Protein transport; Hydrolase; Endoplasmic reticulum;
KW Golgi stack; ATP-binding; Metal-binding; Magnesium; Repeat.
FT NP_BIND 257 264 ATP (POTENTIAL).
FT NP_BIND 537 544 ATP (POTENTIAL).
FT METAL 544 544 MAGNESIUM.
SQ SEQUENCE 742 AA; 81487 MW; 39611058DEAC4A32 CRC64;

Query Match 8.7%; Score 72; DB 1; Length 742;
Best Local Similarity 28.8%; Pred. No. 20;
Matches 34; Conservative 16; Mismatches 44; Indels 24; Gaps 7;

QY 8 FVLIATFGASNDPIA-----DRKIVISDGLVMHER-KPGQEPFYVYV-----MIPKG 57
DB 54 FILSLGCHGSRDGNIALNALQRRHARVSTGDMVSVSRFPENFDLAMLLELEFVKKG 113
QY 58 TEYDQDWILSVGGDHVKKNKPSGRVLYVG---TFP-----YFLTAGAA-VREMDH 106
DB 114 TKSQVDNALST-----OLKRRKTYNQLTVGQKATPEYHGNTVILTVNRADVQGQH 166

RESULT 15
SPR1 ARATH STANDARD; PRT; 303 AA.
AC Q22315; Q22314; Q39201;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pre-mRNA splicing factor SF2 (SR1 protein).
GN SF2 OR A1UG02840 OR F22D16.16 OR F22D16_30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95372342; PubMed=7644475;
RA Lazar G., Schaal T., Maniatis T., Goodman H.M.;
RT "Identification of a plant serine-arginine-rich protein similar to the
RL mammalian splicing factor SF2/ASF."
RN Proc. Natl. Acad. Sci. U.S.A. 92:7672-7676 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Su C.-L., Schuler M.A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gali J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

```

```

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana";
RL Nature 408:816-820 (2000).
CC -!- FUNCTION: Can promote splice site selection in vitro presumably by
CC antagonizing the effects of the Al heterogeneous nuclear
CC ribonucleoprotein. May have an essential function during early
CC plant development.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=SF2A;
CC IsoId=Q22315-1; Sequence=Displayed;
CC Name=2; Synonyms=SF2B;
CC IsoId=Q22315-2; Sequence=VSP_005859, VSP_005860;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Extensively phosphorylated on serine residues in the RS
CC domain.
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M98340; AAA32856.1; -.
DR EMBL; AF001035; AAB71385.1; -.
DR EMBL; AF001035; AAB71386.1; -.
DR EMBL; AC009525; AAF02881.1; -.
DR PIR; F86158; F86158.
DR PIR; S71185; S71185.
DR HSP; P19339; ISXL.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW mRNA processing; mRNA splicing; Nuclear protein; RNA-binding; Repeat;
KW Phosphorylation; Alternative splicing.
FT DOMAIN 7 82 RNA-BINDING (RRM) 1.
FT DOMAIN 96 112 GLY-RICH (HINGE REGION).
FT DOMAIN 119 197 RNA-BINDING (RRM) 2.
FT DOMAIN 198 273 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN 274 303 LYS/PRO/SER-RICH (PSK DOMAIN).
FT VARSPLIC 267 272 SRSRSR -> YGFTYD (in isoform 2).
FT VARSPLIC 273 303 /FTID=VSP_005859.
FT VARSPLIC 273 303 Missing (in isoform 2).
FT CONFLICT 125 125 /FTID=VSP_005860.
FT CONFLICT 127 127 G -> W (IN REF. 1).
FT CONFLICT 139 140 P -> A (IN REF. 1).
FT CONFLICT 140 140 MR -> IA (IN REF. 1).
SQ SEQUENCE 303 AA; 33729 MW; 6530F9CB628B8EFA CRC64;

Query Match 8.7%; Score 71.5; DB 1; Length 303;
Best Local Similarity 25.2%; Pred. No. 8.5;
Matches 26; Conservative 17; Mismatches 51; Indels 9; Gaps 2;

QY 35 DLVWHRKQEQEPFYVYVMPKGT-----YDQDWILSVGGDHVKKNKPSGRVLYV 87
DB 134 DLKDHWRKGDVCFQVYDARDTGTGVDTCTCYEDMKYALKL--DTEFRNFAFNGYVR 191
QY 88 YGTFYFLTAGAAVREMDHFKFTADTGKVDISSKANGHPRSR 130
DB 192 VREYDSRKDSRSPSRSGYSKRSRSGRSGRSRSGRSR 234

```

Thu Sep 9 15:41:32 2004

us-10-768-874-2.rsp

Page 12

Search completed: September 9, 2004, 15:28:36
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:25:00 ; Search time 120 Seconds
(without alignments)
399.656 Million cell updates/sec

Title: US-10-768-874-2
Perfect score: 825
Sequence: 1 MSITQSFVLTALFQAASD.....NWGVMKDGKHYFTVENCQE 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:*
- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvrius:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	825	100.0	152	5	O77009
2	88	10.7	712	2	O85047
3	84	10.2	868	10	Q7XMD8
4	83.5	10.1	389	16	O8A9R8
5	83.5	10.1	474	16	P74698
6	80.5	9.8	280	6	O9TTP1
7	80.5	9.8	329	6	O9TTP2
8	79.5	9.6	1048	2	Q9L8Q9
9	78	9.5	448	16	O8DMZ4
10	77	9.3	501	2	Q9Z4J8
11	77	9.3	501	16	O8DP07
12	77	9.3	557	16	Q97G52
13	76.5	9.3	332	16	Q98RC7
14	76.5	9.3	411	2	O8KXH3
15	76.5	9.3	422	16	Q7V8P5
16	76.5	9.3	768	16	Q92J67

17	76	9.2	185	12	Q8JRZ5
18	76	9.2	492	2	Q9AHT7
19	76	9.2	676	16	Q8NZ52
20	75.5	9.2	245	16	P72948
21	75	9.1	153	2	Q83Y2
22	75	9.1	250	2	Q84B46
23	75	9.1	448	16	Q9KGZ2
24	75	9.1	676	16	Q99XR5
25	75	9.1	676	16	Q8K5M1
26	75	9.1	745	11	Q8ROE8
27	75	9.1	1053	11	Q8VCE0
28	75	9.1	2425	6	Q95MJ0
29	74.5	9.0	204	6	Q9GK7
30	74.5	9.0	341	2	P96948
31	74.5	9.0	1051	2	P70745
32	74	9.0	153	2	Q9S3Y1
33	74	9.0	153	2	Q9R3J6
34	74	9.0	153	2	Q9EXU2
35	74	9.0	360	7	Q98187
36	73.5	8.9	342	16	Q986G7
37	73.5	8.9	350	5	Q95PL6
38	73.5	8.9	578	16	Q8XSG4
39	73.5	8.9	2835	2	Q8G9Q2
40	73	8.8	153	2	Q9S3X9
41	73	8.8	276	2	Q8KTD6
42	73	8.8	468	16	Q83Y38
43	73	8.8	572	16	Q82IP3
44	73	8.8	858	5	Q8IE04
45	73	8.8	870	2	Q03988

ALIGNMENTS

RESULT 1

ID	O77009	PRELIMINARY;	PRT;	152 AA.	
AC	O77009;				
DT	01-NOV-1998	(TREMBlrel. 08, Created)			
DT	01-NOV-1998	(TREMBlrel. 08, Last sequence update)			
DT	01-NOV-1998	(TREMBlrel. 08, Last annotation update)			
DE	Erythema protein SVEP.				
OS	Simulium vittatum (Black fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;				
OC	Simuliidae; Simulium.				
OX	NCBI_TaxID=7192;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Salivary gland secretion;				
RC	MEDLINE=9825234; PubMed=9556538;				
RA	Cupp M., Ribeiro J., Champagne D., Cupp E.;				
RT	"Analyses of cDNA and recombinant protein for a potent vasoactive				
RT	protein in saliva of a blood-feeding black fly, Simulium vittatum.";				
RL	J. Exp. Biol. 201:0-0(0).				
DR	EMBL; U94515; AAC26163.1; --				
SQ	SEQUENCE 152 AA; 17333 MW; 2C8A6CA5FC5F3622 CRC64;				
Query Match 100.0%; Score 825; DB 5; Length 152;					
Best Local Similarity 100.0%; Pred. No. 8.7e-77;					
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1	MSITQSFVLTALFQAASDNP	ADRKRCIVISGDVLV	MHERKPGQFFYYVMIPK	TEY 60
Db	1	MSITQSFVLTALFQAASDNP	ADRKRCIVISGDVLV	MHERKPGQFFYYVMIPK	TEY 60
Qy	61	DDQRWILSVGGHYKLNKFS	GRVLYVCTFDYFLTAGA	AVRMDHFKFTADGTGKYD	IS 120
Db	61	DDQRWILSVGGHYKLNKFS	GRVLYVCTFDYFLTAGA	AVRMDHFKFTADGTGKYD	IS 120
Qy	121	SKANGHPRSRGKNGWVKD	GKHYFTVENCQE		152
Db	121	SKANGHPRSRGKNGWVKD	GKHYFTVENCQE		152

RESULT 2

Q85047 PRELIMINARY; PRT; 712 AA.
 ID O85047
 AC O85047
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Transferrin binding protein B.
 GN TBPB.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Moraxella.
 OX NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3;
 RX MEDLINE=98380363; PubMed=9712766;
 RA Myers L.E., Yang Y.P., Du R.P., Wang Q., Harkness R.E.,
 RA Schryvers A.B., Klein M.H., Loomore S.M.,
 RT "The transferrin binding protein B of moraxella catarrhalis elicits
 RT bactericidal antibodies and is a potential vaccine antigen."
 RL Infect. Immun. 66:4183-4192(1998).
 DR EMBL: AF039311; AAC34274.1;
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0004998; F:transferrin receptor activity; IEA.
 DR InterPro: IPR000437; Prok_lipoprot S.
 DR InterPro: IPR001677; Transferrin_bind.
 DR Pfam: PF01298; Lipoprotein_5; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 SQ SEQUENCE 712 AA; 76863 MW; C50808DD11AF2C09 CRC64;

Query Match 10.7%; Score 88; DB 2; Length 712;
 Best Local Similarity 27.8%; Pred. No. 3.2;
 Matches 37; Conservative 10; Mismatches 44; Indels 42; Gaps 6;

QY 38 MHERKPGFFPYVYVIMPKGTEYDQWILESVGGDHKL-----KNKFSGR 84
 DB 276 WTDARKGOSFSF-----ERRAGRYSAMSHYPSLLTDDKNKPNY 318
 QY 85 YLVYVG-----TDY-----FLTAGAAREMDHFKFTADGTGYDISSKANGHPRSGKNWGV 136
 DB 319 NDEYGHSSFEFTDFSKSLTGLFSLQDHKGKVKTKRYDINARIHGN-RFRGSATAI 377
 QY 137 MKDGE---KHFT 146
 DB 378 NKDNESKAKHPFT 390

RESULT 3

Q7XMD8 PRELIMINARY; PRT; 868 AA.
 ID Q7XMD8
 AC Q7XMD8
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE OSUNFA0061G20.21 protein.
 GN OSUNFA0061G20.21.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
 RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
 RA Ren S.X., Lv G., Lin W., Gu W.G., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
 RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,

RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
 RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL662974; CA604685.1;
 SQ SEQUENCE 868 AA; 98333 MW; 28BD67842B4EB651 CRC64;

Query Match 10.2%; Score 84; DB 10; Length 868;
 Best Local Similarity 24.8%; Pred. No. 11;
 Matches 33; Conservative 20; Mismatches 40; Indels 40; Gaps 7;

QY 43 PQEFPYVYVIMPKGTEYDQWIL-----ESVGGDH-----KLKN 79
 DB 139 PAEAKNYTFTLPAGTCKVKRWTLKKWAEQFSFGKGLYQKYLKGTGTPNFTFPKLRD 198
 QY 80 KPSGRYLVGTFDYFLTAGAARE-----MDHFKFTADGTGYD-----ISSKANGHP 127
 DB 199 HWD-EFVAYKTADGSLVFGDQIREAVRLTDAVEASSQGTFRPDRERDELSALQTPPEH 257
 QY 128 -RSRGKN---WGV 136
 DB 258 GRTRGKGVIPWKI 270

RESULT 4

Q8A9R8 PRELIMINARY; PRT; 389 AA.
 ID Q8A9R8
 AC Q8A9R8
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative phosphatidylinositol-4-phosphate 5-kinase.
 GN BT0747.
 OS Bacteroides thetaiotaomicron.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_TaxID=818;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI-5482 / ATCC 29148;
 RX MEDLINE=24550858; PubMed=12663928;
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
 RA Chiang H.C., Hooper L.V., Gordon J.I.;
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
 RL Science 299:2074-2076(2003).
 DR EMBL: AE016929; AA075854.1;
 DR GO: GO:0016301; P:kinase activity; IEA.
 DR InterPro: IPR003409; MORN.
 DR Pfam: PF02493; MORN; 14.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 389 AA; 44241 MW; CE805074D37E851B CRC64;

Query Match 10.1%; Score 83.5; DB 16; Length 389;
 Best Local Similarity 24.4%; Pred. No. 4.2;
 Matches 33; Conservative 20; Mismatches 59; Indels 23; Gaps 7;

QY 30 VLSGDGLVHHRKPGQEPYVYVIMPKGTEYDQWILESVGGD---HYLKNKFSGRYL 86
 DB 72 VFKNEDVYEGEYIKKRGYGLYSPDGKEYGQ-WFQDQOQKGYIYFMNNRYDGMWF 130
 QY 87 V-----YGTFFYFLTAGAAREMDHFKFTADGTGYDISSKANGH---PRSRGK---NW- 134
 DB 131 QDYQHAGTGM-YHNGDLYVGNWANDKREGEQTYTWANGAKYSCHWKNDKNGKGTMMWD 189
 QY 135 -----GVMKDGEKH 143
 DB 190 DGCKYDGDWKKDVRH 204

RESULT 5

P74698 PRELIMINARY; PRT; 474 AA.
 ID P74698
 AC P74698
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DR InterPro: IPR006425; Glucan glucosid.
DR InterPro: IPR000165; Glyco_hydro_15.
DR InterPro: IPR008928; Glyco_trans_5np.
DR Pfam: PF00723; Glyco_hydro_15; 1.
DR TIGRfam: TIGR01535; Glucan glucosid; 1.
DR PROSITE: PS00820; GLUCOAMYLASE; 1.
SQ SEQUENCE 1048 AA; 109140 MW; 09D4D124B478DF4F CRC64;

Query Match 9.6%; Score 79.5; DB 2; Length 1048;

Best Local Similarity 23.9%; Pred. No. 40;
Matches 38; Conservative 16; Mismatches 60; Indels 45; Gaps 8;

QY 8 FVLTLAIFCAASDN-----PIARRKCVISDGLVWHERKPGQFPYVYVMPKGT 58
DB 396 PPTSRVDTGTCQNGIQIDETAFFILLANOIGRTDAGFYRELKPAAD-----YLAAGP 450
QY 59 EYDDORWILESGVGHYKL-----KNKFSGRYLVYGTDFYFLTAGAA 100
DB 451 KTPQERW--EETGG--YSTSLASQIALAALAAADIAGKGDAGSAVYRA-----TADFW 501
QY 101 VREMDHFXTAD--GTGKYDISSKANGHPRSRG--KNWG 135
DB 502 QRSTKWMFTTNGPVGDGKYLRISATGNPDGATRDWG 540

RESULT 9

Q8DMZ4 ID Q8DMZ4 PRELIMINARY; PRT; 448 AA.

AC Q8DMZ4; 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE Choline binding protein D.
GN CBPD OR SFR2006.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1] SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=1154234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL: AF008565; AAL00808.1; --
DR PIR: C98122; C98122.
DR InterPro: IPR007921; CHAP.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003646; SH3_bac.
DR Pfam: PF05257; CHAP; 1.
DR Pfam: PF01473; CW_binding_1; 3.
DR SMART: SM00287; SH3b; 2.
DR PROSITE: PS00911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 448 AA; 50396 MW; C136BE40CB530381 CRC64;

Query Match 9.5%; Score 78; DB 16; Length 448;

Best Local Similarity 23.8%; Pred. No. 19;
Matches 41; Conservative 19; Mismatches 38; Indels 74; Gaps 11;

QY 43 PGQEPFYVYVMPKGTGYDDQRW-----ILSVGGDHY- 75
DB 217 PGEKHYDQIL-----EKDGKWLSTAYNGSVRYVQLEAVNKPNLGNVLSLTGTGTHYF 271
QY 76 KLNKFSGRYLVYGT-FDYFLTAGAA-----VREMDHFK---FTA-DGTGKY----- 117

Db 272 KIKSAIKTEPLVSATVIDYYPGEKHYDQILEKDGKWLSTAYNGSVRYVQLEGVTS 331
QY 118 -----DISS-----KANG---HPRSRG--KNWGVKDGKHYF 145
DB 332 QNYQNSGNISSYSGNNSSTVGVKKINGSWYHFKSNGSKSGTWLKDSSWY 383

RESULT 10

Q9Z4J8 ID Q9Z4J8 PRELIMINARY; PRT; 501 AA.

AC Q9Z4J8; 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE 1,4-beta-N-acetylmuramidase precursor (EC 3.2.1.17).
GN LYTC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1] SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=W31;
RX MEDLINE=99340545; PubMed=10411730;
RA Garcia P., Gonzalez M.P., Garcia B., Garcia J.L., Lopez R.;
RT "The molecular characterization of the first autolytic lysozyme of
Streptococcus pneumoniae reveals evolutionary mobile domains";
RL Mol. Microbiol. 33:128-138(1999).
DR EMBL: AJ009639; CA08765.1; --
DR GO: GO:0016798; F:hydrolase activity; IEA.
DR GO: GO:0003796; F:lysozyme activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding_1; 5.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 501 1,4-BETA-N-ACETYLMURAMIDASE.
SQ SEQUENCE 501 AA; 58682 MW; A6774B6DBF2EB704 CRC64;

Query Match 9.3%; Score 77; DB 2; Length 501;

Best Local Similarity 25.0%; Pred. No. 27;
Matches 31; Conservative 18; Mismatches 43; Indels 32; Gaps 8;

QY 34 GDLVWHERKPGQEPFYVYVMPKGTGYDDQRWILSVGGDHYKLR-----NKFSGRYL 86
DB 186 GMIQNEWLYDPAYSAYFLKSDGT-YANQEW--QKVGKWWYFKKNGWMAENWQGNV- 241
QY 87 VYGTDFYFLTAGAAVRE---MD--HFKFTADGTGKYDISSKANGHPRSRGKNWGVKQGE 141
DB 242 -----YLTGGAMATDEIVMDGARYIFAASGELK-----EKKDLNVGV-VHRDCK 285
QY 142 KHYF 145
DB 286 RYFF 289

RESULT 11

Q8DP07 ID Q8DP07 PRELIMINARY; PRT; 501 AA.

AC Q8DP07; 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE 1,4-beta-N-acetylmuramidase (EC 3.2.1.17).
GN LYTC OR SP1431.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1] SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;

RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Galmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'gara M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
 RA Glass J.I.,
 RT "Genome of the bacterium *Streptococcus pneumoniae* strain R6.";
 RL J. Bacteriol 193:5709-5717(2001).
 DR EMEL; ABX08513; AAL00235.1; - -
 DR FIR; F98050; F98050.
 DR GO; GO:0016798; F:Hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0003796; F:lysozyme activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro: IPR002479; CW binding.
 DR Pfam: PF01473; CW_binding_1; 5.
 KW Glycosidase; HydroLase; Complete proteome.
 SQ SEQUENCE 501 AA; 58681 MW; A2D74FC911241DA4 CRC64;

Query Match	9.3%;	Score 77;	DB 16;	Length 501;	
Best Local Similarity	25.0%;	Pred. No. 27;			
Matches	31;	Conservative 18;	Mismatches 43;	Indels 32;	Gaps 8;
QY	34	GDLMHERKQGEPPYYVYMLPKGTETDDQEWILES VGGDHYK LK-----NKFSGRYL	86		
Db	186	GAMIQENLWLDPAYSAYFLKSDGT-YANQEW--QKVGGRWYFPKKGWYMARNEWQGN Y- 241			
QY	87	VYGTFTDYFLTAGAAVRE---MD--HFKFTADGTGKYDISSKANGHPSPRGKNWGVMKDGE	141		
Db	242	-----YLTGSGAMATDEVINDGARYIFAASGELK-----EKXDLNVGV-VHRDGG	285		
QY	142	KHYF	145		
Db	286	RYFF	289		

RESULT 12	Q97G52	PRELIMINARY;	PRT;	557 AA.
ID	Q97G52			
AC	Q97G52;			
DT	01-OCT-2001 (TrEMBLrel. 18, Created)			
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Extracellular neutral metalloprotease, NPPE.			
GN	CAC2517.			
OS	Clostridium acetobutylicum.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1488;			
OX	[1]			
RN	SEQUENCE FROM RN A.			
RP	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;			
RC	MEDLINE=21359325; PubMed=11466286;			
RY	Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.			
RA	Gibson R.L., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,			
RA	Tatousov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.S.			
RA	Bennett G.N., Koonin E.V., Smith D.R.;			
RA	"Genome sequence and comparative analysis of the solvent-producing			
RT	bacterium Clostridium acetobutylicum."			
RL	J. Bacteriol. 183:4823-4838 (2001).			
DR	EMBL; AR007750; AAK80471.1; -.			
DR	PIR; A60152; A60152.			
DR	PIR; D97210; D97210.			
DR	MEROPS; M04.011; -.			
DR	GO; GO:0005576; C:extracellular; IEA.			
DR	GO; GO:0004222; F:metalloendopeptidase activity; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	GO; GO:0005508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR001570; Peptidase M4.			
DR	InterPro; IPR005075; Peptidase M4 N.			
DR	InterPro; IPR006025; Pept M Zn BS.			

```

DR Pfam; PF01447; Peptidase_M4; 1.
DR Pfam; PF02868; Peptidase_M4_C; 1.
DR Pfam; PF03413; Pep_M4_propep; 1.
DR PRINTS; PR00730; THERMOLYSIN
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Protease; Metalloprotease; Complete proteome.
SQ SEQUENCE 557 AA; 60362 MW; DBBB1854F494669 CRC64;

Query Match          9.3%; Score 77; DB 16; Length 557;
Best Local Similarity 23.4%; Pred. No. 32;
Matches 36; Conservative 18; Mismatches 62; Indels 38; Gaps 7;

Qy 9 VLTLAFGAASDNPIDARKCIVISDGLVMHRRKPGQEPFY-----VYMI 54
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 336 IISTAHVGSQDYDAYWDGSGMVYGDGD-----GSEFTYFGDLDVVGHELTHGVQY 387

Qy 55 PKGTEYDDQRWIL-ESVGG-----DHYLKNKFSGRYLYVGTGYFLTACAAVREM 104
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 388 TANLINVEDQSALNESGSDVFGVLITQDYKNYKN---GGNWAIFSADSVVVGDLFLNLT 444

Qy 105 DH--FKFETADGTGKYDISSKANGHPRSRGKNQGV 136
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 445 THRALRSLANPT-LYDQPDNNNNYVNTSDDNGGV 477

```

RESULT 13	
Q98RC7	PRELIMINARY; PRT; 332 AA.
ID	Q98RC7
AC	Q98RC7
DT	01-OCT-2001 (TrEMBLrel. 18, Created)
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Restriction-modification enzyme subunit M3 (EC 2.1.1.72).
DE	MYPU 0820.
GN	Mycoplasma pulmonis.
OS	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC	NCBI_TaxID=2107;
NC	[1]__SEQUENCE FROM N.A.
RN	STRAIN=UAB CT1P;
RP	MEDLINE=21267165; PubMed=11353084;
RX	Chambaud I., Heilig R., Ferris S., Barbe V., Sanson D., Galisson
RA	Mosser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA	Blanchard A.;
RA	"The complete genome sequence of the murine respiratory pathogen
RT	Mycoplasma pulmonis.";
RL	Nucleic Acids Res. 29:2145-2153(2001).
RT	EMBL; AL445563; CAC13255.1; -.
DR	PIR; B90522; B90522.
DR	MyPulist; MYPU 0820. -.
DR	GO; GO:0003677; F:DNA binding; IEA.
DR	GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR	GO; GO:0009007; F:site-specific DNA-methyltransferase (adenin. .
DR	GO; GO:0016740; F:transferase activity; IEA.
DR	GO; GO:0006306; P:DNA methylation; IEA.
DR	GO; GO:0006304; P:DNA modification; IEA.
DR	InterPro; IPR003665; Methylase M.
DR	InterPro; IPR002296; N12N6_mtfase.
DR	InterPro; IPR003356; N6 DNA Mtase.
DR	InterPro; IPR002052; N6 Mtase.
DR	Pfam; PF02506; Methylase M; 1.
DR	Pfam; PF02384; N6 Mtase_1.
DR	PRINTS; PR00507; N12N6MTFRASE.
DR	PROSITE; PS00092; N6_MTASE; 1.
DR	Transferrase; Methyltransferase; Complete proteome.
DR	SEQUENCE 332 AA; 38059 MW; PAAIF85CA1FE140 CRC64;

Query Match 9.3%; Score 76.5; DB 16; Length 332;
Best Local Similarity 23.0%; Pred. No. 18;
Matches 35; Conservative 16; Mismatches 60; Indels 41; Gaps 6;

QY 10 LTLATFGASDNPIDRKC-----IVISDGLVNHKKPGCEFPYYVMIPKGTETDDQRW 65
 : : :
 : : :

OS Prochlorococcus marinus (strain MIT 9313).
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=74547;
 RN [1]

RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Ahlgr n N.A., Arellano A., Coleman M., Hauser L., Hess W.R.

Shaw S.-I., Zinsler E.C., Goffard D., and G.S. (1979) Molecular
RA RA
RA RA
Webb E.A., Zinsler E.C., Chisholm S.W.;
RT "Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic
RT niche differentiation."
RT

DR	EMBL; BX572095; CAE20465.1; -	
RD	Complete proteome.	
SQ	SEQUENCE 422 AA; 47864 MW; DA22A25063A33E74 CRC64;	

Query Match	9.3%;	Score 76.5;	DB 16;	Length 422;
Best Local Similarity	32.1%;	Pred. No. 25;		
Matches 26;	Conservative 10;	Mismatches 32;	Indels 13;	Gaps 3;

Qy	35	DLVHERKPGQEFPPYVYVMTLPGKTEYDDQRMLESVGGSDHVKLKNKFSGRYLIVYGTDFY	94
Db	309	DLAHLRKINGD-QDYVFWSPGR---SKPHIDESALNHLKVLRLVYGGRTAHGLRSLP	364

Qy	95	LTAGAAV-----REMDH	106
Db	365	LTAGQVLKFSAEVIQRQMDH	385

Search completed: September 9, 2004, 15:31:28
Job time : 123 secs

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:06 11 November 2014

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:14:53 ; Search time 122 Seconds
(without alignments)
352.026 Million cell updates/sec

Title: US-10-768-874-2

Perfect score: 825

Sequence: 1 MSITQSPFVLTALFQAASD.....NWGVKMDGKHFTVENCQE 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	825	100.0	152	2	AAW74589
2	88	10.7	722	2	AAW74589 Simulium
3	84.5	10.2	1297	7	AAY43379 M. catarr
4	80.5	9.8	280	2	ADG94925 E. faeciu
5	80.5	9.8	329	2	AAY41078 Canine B7
6	79.5	9.6	1047	3	AAY41076 Canine B7
7	78	9.5	1129	6	AAB35740 Glucodext
8	76.5	9.3	332	7	ABR52906 Protein s
9	76.5	9.3	448	3	ADG23977 Protein s
10	76	9.2	471	2	AAY71551 Florida b
11	76	9.2	471	5	AAW61216 Streptoco
12	76	9.2	471	5	APP54635 S. pneumo
13	76	9.2	490	7	ADG45239 S. pneumo
14	76	9.2	506	6	AAY81745 Streptoco
15	75.5	9.2	179	3	ABU02034 S. pneumo
16	75.5	9.2	180	3	ABU02034 S. pneumo
17	75	9.1	446	5	AGL10251 Arabidops
18	75	9.1	446	5	AAU76561 Rat hist
19	75	9.1	448	5	AAU76564 Human his
20	75	9.1	448	6	AAY71285 Streptoco
21	75	9.1	676	5	ABU02732 S. pneumo
22	75	9.1	676	5	ABP29594 Streptoco
23	75	9.1	1013	7	ABU46865 Protein e
24	75	9.1	1013	7	ADB85134 Rat NA,K-
25	75	9.1	1013	7	ADG55704 Human Pro
					ADG55722 Human Pro

26	75	9.1	1013	7	ADG55724 Rat Prote
27	75	9.1	1013	7	ADG55707 Human Pro
28	75	9.1	1013	7	ADG55720 Rat Prote
29	75	9.1	1014	7	ADG55706 Rat Prote
30	75	9.1	1014	7	ADG55703 Rat Prote
31	74.5	9.0	837	6	ABP77467 N. gonorr
32	74.5	9.0	878	6	ABU37086 Protein e
33	74.5	9.0	1051	2	AAW52304 Glucodext
34	73.5	8.9	367	5	ABU08211 Escherich
35	73.5	8.9	2835	5	ABR98574 Dextran s
36	73.5	8.9	2835	6	ABR55594 Amino aci
37	73	8.8	870	2	AAI14529 Mosquitot
38	73	8.8	870	2	AAI14529 Mosquitot
39	73	8.8	870	2	AAI14529 Mosquitot
40	72.5	8.8	110	3	AAW75410 B. sphaer
41	72.5	8.8	110	3	AAW75410 B. sphaer
42	72.5	8.8	110	3	AAW75410 B. sphaer
43	72.5	8.8	162	3	AAW75410 B. sphaer
44	72.5	8.8	162	3	AAW75410 B. sphaer
45	72.5	8.8	164	3	AAW75410 B. sphaer

ALIGNMENTS

RESULT 1

AAW74589
ID AAW74589 standard; protein; 152 AA.

XX AC AAW74589;

XX XX

DT 25-MAR-2003 (revised)

DT 09-DEC-1998 (first entry)

XX XX

DE Simulium vittatum salivary gland erythema protein (SVEP).

XX XX

XX Blood-feeding black fly; salivary gland erythema protein; SVEP;

KW vasodilative activity; wound healing; suppression; immune response;

KW treatment; peripheral resistance; atherosclerosis; heart failure;

KW hypertension; graft rejection; autoimmune disease.

XX OS Simulium vittatum.

XX OS

PN WO9840089-A1.

XX XX

ED 17-SEP-1998.

XX XX

PF 12-MAR-1998; 98WO-US004795.

XX XX

PR 13-MAR-1997; 97US-0040418P.

XX XX

PR 06-MAR-1998; 98US-00036355.

XX XX

PA (AUBU) UNIV AUBURN.

XX XX

PA (UYAR-) UNIV ARIZONA.

XX XX

PI Cupp MS, Ribeiro JMC, Cupp EW, Swaim SP;

XX XX

DR WPI; 1998-506475/43.

XX XX

DR N-PSDB; AAV54348.

XX XX

PS Claim 5; Page 18; 27pp; English.

XX XX

CC The present sequence represents Simulium vittatum (blood-feeding black

CC fly) salivary gland erythema protein (SVEP). The proteins exhibit

CC vasodilative activity and wound healing promoting properties, as well as

CC the capacity to suppress certain immune responses in a mammal. The

CC vasodilative or immunomodulating proteins from the salivary glands of

CC Simulium species may be administered to a mammal in order to treat

CC peripheral resistance or to modulate the immune response. The proteins

New isolated vasodilative protein from salivary glands of Simulium species - useful, e.g. for regulating blood pressure or promoting wound healing and as immunomodulator.

CC can also be used for treating atherosclerosis of extremities, heart
CC failure, hypertension, peripheral resistance, stenoses and in particular
CC peripheral vasodilation. These proteins may also be used to suppress the
CC immune system, to inhibit or prevent the development of antibodies or
CC cellular immunity to a protein, treat graft rejection and autoimmune
CC diseases. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-
CC MAR-2003 to correct PI field.)
XX
SQ Sequence 152 AA;

Query Match 100.0%; Score 825; DB 2; Length 152;
Best Local Similarity 100.0%; Pred. No. 8.1e-87;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSITQFFVLTALFAGASDNFIADKCIIVISDGLVMHERKPGQFFYYVMIKGTET 60
DB 1 MSITQFFVLTALFAGASDNFIADKCIIVISDGLVMHERKPGQFFYYVMIKGTET 60
QY 61 DDQWILSVGGDHYKLNKFGRLVYGTFFYFLTAGAAVREMDHFKFTADGTGKYDIS 120
DB 61 DDQWILSVGGDHYKLNKFGRLVYGTFFYFLTAGAAVREMDHFKFTADGTGKYDIS 120
QY 121 SXANGHPRSRGKNWGMKDGKHYFTVENCQE 152
DB 121 SXANGHPRSRGKNWGMKDGKHYFTVENCQE 152

RESULT 2
AAY43379
ID AAY43379 standard; protein; 722 AA.
XX
AC AAY43379;
DT 26-JAN-2000 (first entry)
XX
DE M. catarrhalis strain 3 tbp2 protein.
XX
KW TbpB Gene; Tbp2; transferrin binding protein; diagnosis; otitis media;
KW genetic immunisation; Moraxella infection; antigen; vaccine; detection;
KW antitumour antibody production; therapy.
OS
OS Moraxella catarrhalis.
XX
PN WO9522947-A2.
XX
PD 21-OCT-1999.
XX
PF 12-APR-1999; 99WO-CA000307.
XX
PR 14-APR-1998; 98US-00059584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Myers LE, Schryvers AB, Harkness RE, Loosmore SM, Du R, Yang Y;
PI Klein MH;
XX
WPI; 1999-620376/53.
DR N-PSDB; AAZ31947.
XX
XX Nucleic acid encoding transferrin binding protein 2 of Moraxella
PT catarrhalis, useful for diagnostics, immunization and recombinant protein
PT production.
XX
PS Claim 7; Fig 4; 114pp; English.

CC This sequence is the Moraxella catarrhalis strain 3 transferrin binding
CC protein (Tbp2) of the invention. The DNA sequence is also referred to as
CC the TbpB gene. The TbpB gene is used to produce recombinant Tbp2; for
CC identification or diagnosis of Moraxella, or for cloning related species,
CC using hybridisation assays; and for genetic immunisation against
CC Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as
CC antigens, either in vaccines (including components of conjugate vaccines
CC that contain antigens from other bacteria or from tumours, in which case

CC they elicit production of antitumour antibodies that may be coupled to
CC chemotherapeutic agents or biologically active agents) or to raise
CC antibodies (for use as diagnostic reagents and for treating Moraxella
CC infections), also for detecting Moraxella antibodies
XX
SQ Sequence 722 AA;

Query Match 10.7%; Score 88; DB 2; Length 722;
Best Local Similarity 27.8%; Pred. No. 0.81;
Matches 37; Conservative 10; Mismatches 44; Indels 42; Gaps 6;
QY 38 MHERKPGQFFYYVMIKGTETDQRMILSVGGDHYKL-----KNKFSGR 84
DB 286 MTDKKGQSPSF-----ERRADRYSAMSSHEVPSLLTDDKNKPDNY 328
QY 85 YLVYG-----TFDY---FLTAGAAVREMDHFKFTADGTGKYDISKANGHPRSRGKNWGV 136
DB 329 NDEYGHSEFTVDFSKSLTGLFSLNLDHKGKVTTKRYDINARIHGN-RFRGSATAI 387
QY 137 MKDGF---KHVFT 146
DB 388 MKDNESKAKHFT 400

RESULT 3
ADC94925
ID ADC94925 standard; protein; 1297 AA.
XX
AC ADC94925;
DT 01-JAN-2004 (first entry)
XX
DE E. faecium protein sequence SEQ ID 4552.
XX
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2003-799836/75.
DR N-PSDB; ADC91271.
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
PS Example 1; SEQ ID NO 4552; 243pp; English.
XX
CC The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids is useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract

CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.

XX SQ Sequence 1297 AA;
 Query Match 10.2%; Score 84.5; DB 7; Length 1297;
 Best Local Similarity 28.8%; Pred. No. 4.6;
 Matches 36; Conservative 14; Mismatches 36; Indels 39; Gaps 8;
 QY 30 VISDGLVMHE-----RKQGEPPYV-YMI-----PKTFYDDQR--WIL-----ES 69
 DB 920 VIVDGEKVIHPICEPDSFVYKDSSETTYNAEYKIIESQNPGEFKPEPNDTWILRLYDNA 979
 QY 70 VGGDHVYKLNKPSGRYL-----VYGTDFYFLTAGAAVREMDHPKFTADGT 114
 DB 980 IGTIQLKQTGEENWQYLTVTTBEDRLWVAFYATNTFDY---REIRVRKLDHGHGTLDDGA 1036
 QY 115 GKYDI 119
 DB 1037 G-FDI 1040

RESULT 4
 ID AAY41078
 XX AAY41078 standard; protein; 280 AA.
 AC AAY41078;
 XX
 XX 20-DEC-1999 (first entry)
 DT
 XX Canine B7-2S protein.
 DE
 XX B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 KW allergic reaction; infectious disease; tumor development; canine;
 KW graft rejection; inflammation; arthritis; atopic dermatitis.
 XX
 XX Canis familiaris.
 OS
 XX
 XX WO9947558-A2.
 XX
 XX 23-SEP-1999.
 XX
 XX 19-MAR-1999; 99WO-US006187.
 XX
 XX 19-MAR-1998; 98US-0078765P.
 PR 17-APR-1998; 98US-00062597.
 XX
 XX (HESK-) HESKA CORP.
 PA
 XX
 XX Sim G, Yang S, Sellins KS;
 PI
 XX WPI; 1999-571822/48.
 DR N-PSDB; AAZ27921, AAZ27923.
 XX
 XX New isolated B7 and CTLA4 nucleic acids, used to develop products for
 PT treating, e.g. autoimmune and atopic diseases.
 PT
 XX Claim 4; Page 109-111; 148pp; English.

XX The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritis and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening

XX SQ Sequence 280 AA;
 Query Match 9.8%; Score 80.5; DB 2; Length 280;
 Best Local Similarity 27.1%; Pred. No. 1.6;
 Matches 29; Conservative 13; Mismatches 38; Indels 27; Gaps 4;
 QY 1 MSITQSFFVLTALFAGAA-----SDNPIADRKCIIVISDGD-LVMH 39
 DB 7 MELNNILFVMTLLLYGAAMKSOAYFNKGTGELPCHFTNSQNSISLDELVVFWDQDKLVLY 66
 QY 40 ERKPGQEPFYVYMIIPKG-TEYDDQRWILES VGGDHVYKLNKPSGRY 85
 DB 67 ELYRGKENPQVHRKYKGRTSFDDKNWTLRL-----HNIQIKDKGLY 108

RESULT 5
 ID AAY41076
 XX AAY41076 standard; protein; 329 AA.
 AC AAY41076;
 XX
 XX 20-DEC-1999 (first entry)
 DT
 XX Canine B7-2 protein.
 DE
 XX B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 KW allergic reaction; infectious disease; tumor development; canine;
 KW graft rejection; inflammation; arthritis; atopic dermatitis.
 XX
 XX Canis familiaris.
 OS
 XX
 XX WO9947558-A2.
 XX
 XX 23-SEP-1999.
 XX
 XX 19-MAR-1999; 99WO-US006187.
 XX
 XX 19-MAR-1998; 98US-0078765P.
 PR 17-APR-1998; 98US-00062597.
 XX
 XX (HESK-) HESKA CORP.
 PA
 XX
 XX Sim G, Yang S, Sellins KS;
 PI
 XX WPI; 1999-571822/48.
 DR N-PSDB; AAZ27913, AAZ27915.
 XX
 XX New isolated B7 and CTLA4 nucleic acids, used to develop products for
 PT treating, e.g. autoimmune and atopic diseases.
 PT
 XX Claim 4; Page 97-99; 148pp; English.

XX The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritis and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening

XX SQ Sequence 329 AA;
 Query Match 9.8%; Score 80.5; DB 2; Length 329;
 Best Local Similarity 27.1%; Pred. No. 2;
 Matches 29; Conservative 13; Mismatches 38; Indels 27; Gaps 4;
 QY 1 MSITQSFFVLTALFAGAA-----SDNPIADRKCIIVISDGD-LVMH 39
 DB 7 MELNNILFVMTLLLYGAAMKSOAYFNKGTGELPCHFTNSQNSISLDELVVFWDQDKLVLY 66

XX	Protein sequence #SEQ ID 677
DE	

W

```

OS Unidentified.
XX WO2003000840-A2.
XX 03-JAN-2003.
XX 15-MAY-2002; 2002WO-US015983.
XX 21-JUN-2001; 2001US-0300189P.
XX 30-JUL-2001; 2001US-0309006P.
XX 22-JAN-2002; 2002US-0351336P.
XX (DIVE-) DIVERSA CORP.
XX (MADD/) MADDEN D.
XX Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;
XX Short JM, Burk M;
XX WPI; 2003-201417/19.
XX N-PSDB; ADC23976.
XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-
XX 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl
XX lactic acid derivative and for producing pharmaceutical composition, and
XX food additive.
XX Claim 40; SEQ ID NO 244; 560pp; English.
XX This invention relates to nitrilases and the nucleic acids that encode
XX these enzymes thereof. Specifically, it refers to polypeptides that
XX exhibit nitrilase activity, i.e. the ability to directly hydrolyse
XX nitriles or cyanohydrins into their corresponding carboxylic acids and
XX ammonia. Nitrilases have commercial utility as biocatalysts for use in
XX the synthesis of enantiomerically pure aromatic and aliphatic amino
XX acids, as well as hydroxy acids, which are important for the development
XX of chiral medicines. Furthermore, the present invention describes
XX nitrilases, isolated from mesophilic microorganisms, that have improved
XX activity and stability at increased pH and temperature. They are also
XX inexpensive, efficient catalysts, have broad substrate specificity and
XX are capable of chiral differentiation. This polypeptide is a protein
XX sequence that exhibits nitrilase activity of the invention.
XX Sequence 332 AA;
XX
XX Query Match 9.3%; Score 76.5; DB 7; Length 332;
XX Best Local Similarity 25.0%; Pred. No. 5.9;
XX Matches 32; Conservative 17; Mismatches 50; Indels 29; Gaps 6;
XX
XX 18 ASDNPIADKICIVISDGLVMHKKPGQEPFYVYVIMPKGTEYDDQWILESD---VGGD 73
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 205 ASRYAFEGRCFVLAAG-LMKVVDIPPELE-----LPSQMSRESEDLLEGGSAVIGPD 258
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 74 HYKLNKFGSYLYVGTDP--YFUTAGAAVREMDHFKFTADGTGKY-----DISKA 123
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 259 -----GKYIVFPLFDREAILTADLEALACDRKMTLDTVGHYRPRDLHLEFRKQQ 309
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 124 NGHPRSRG 131
XX : : : : :
XX 310 SGHIAGAG 317
XX
XX RESULT 9
XX AA71551
XX ID AA71551 standard; protein; 448 AA.
XX AC
XX AA71551;
XX 12-OCT-2000 (first entry)
XX DE Florida bitterbush delta-6 fatty acid desaturase.
XX KW Florida bitterbush; delta-6 fatty acid desaturase; tariric acid;
XX transgenic plant; fatty acid; membrane-bound desaturase.

```

```

XX OS
XX OS Picramnia pentandra.
XX PN WO200032790-A2.
XX XX
XX PD 08-JUN-2000.
XX XX
XX PF 02-DEC-1999; 99WO-US028589.
XX XX
XX PR 03-DEC-1998; 98US-0110784P.
XX XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO 'E I.
XX XX
XX PI Cahoon EB, Cahoon RE, Hitz WD, Kinney AJ;
XX XX
XX DR WPI; 2000-412336/35.
XX DR N-PSDB; AAD01349.
XX XX
XX PT Polynucleotide encoding delta-6 desaturase enzyme useful for producing
XX PT transgenic plants and for producing antibodies specific to which is
XX PT useful for screening cDNA expression libraries.
XX PS Claim 10; Page 40-41; 57pp; English.
XX CC
XX CC The present sequence is a delta-6 fatty acid desaturase protein sequence
XX CC from clone pps.pk0011.d5.fis isolated from Florida bitterbush developing
XX CC seed cDNA library, pps. The delta-6 desaturase enzyme catalyses the
XX CC formation of tariric acid, a fatty acid that has a triple bond at the
XX CC delta-6 carbon. The present sequence is useful for producing transgenic
XX CC plants having altered levels of delta-6 desaturase which in turn would
XX CC alter the fatty acid composition. The enzyme is also useful for producing
XX CC polyclonal or monoclonal antibodies. The polynucleotide is useful as
XX CC primer or probe for screening cDNA libraries to isolate desired full-
XX CC length cDNA clones
XX CC
XX SQ Sequence 448 AA;
XX
XX Query Match 9.3%; Score 76.5; DB 3; Length 448;
XX Best Local Similarity 28.6%; Pred. No. 8.9;
XX Matches 33; Conservative 17; Mismatches 46; Indels 17; Gaps 7;
XX
XX 31 ISDGLVMHKKPGQEPFYVYVIMPKGTEYDDQWILESDGHDYKLNKFGSYLYVGT 90
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 8 ISQADLAKH-KQPGD-----LWISIKGVYDISKWTKEHFGGELPLL--SPAGQDVT--- 56
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 91 FDYPLT--AGAAVREMDHFKFTADGTGKYDISKANGHPR--SEGKNGWVK 138
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 57 -DAFIAYHPGTAWQYLDLF-FTGYVQDYGVSEMSKDYRLRLVSEFSTKMLFK 106
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 10
XX AA61216
XX ID AA61216 standard; protein; 471 AA.
XX XX
XX AC AA61216;
XX XX
XX DT 02-OCT-1998 (first entry)
XX XX
XX DE Streptococcus pneumoniae SP0091 protein.
XX XX
XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KW detection; pneumonia; otitis media; meningitis.
XX XX
XX OS Streptococcus pneumoniae.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 164
XX FT /note= "encoded by CTSC"
XX XX
XX PN W09818930-A2.
XX XX
XX PD 07-MAY-1998.
XX XX

```

PF 30-OCT-1997; 97WO-US019422.
 XX
 PR 31-OCT-1996; 96US-0029960P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
 XX
 PI WPI; 1998-272224/24.
 DR N-PSDB; AAV27402.
 XX

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
 PT - or their epitope-containing fragments, useful in protective or
 PT therapeutic vaccines, and for diagnosis.
 XX

PS Claim 11; Page 81; 118pp; English.
 XX

XX The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose
 XX

XX Sequence 471 AA;
 SQ

Query Match 9.2%; Score 76; DB 2; Length 471;
 Best Local Similarity 25.2%; Pred. No. 11;

Matches 30; Conservative 19; Mismatches 48; Indels 22; Gaps 6;

QY 34 GDLVHERKPGQEPFYVYVIMPKGTEDDQWILESVGGDHYKLGK-----NKFSGRYL 86
 DB 156 GAMQNEWLYDPAYGAYFLKSDGT-YANQEW--QKVGKWIYFKKGMARNEWQG--- 209

QY 87 VYGTDFYLTAGAAVREMDHFKFTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYF 145
 DB 210 -----NYLTGSGA---MATDEVIMDGTTRYIFAASGELKEKDLNVGV-VHRDGRKRYF 259

RESULT 11
 ABP54635
 ID ABP54635 standard; protein; 471 AA.
 XX
 AC ABP54635;
 XX

DT 04-SEP-2002 (first entry)
 XX

DE S. pneumoniae SP091 protein sequence SEQ ID NO:158.
 XX

KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection.
 XX

OS Streptococcus pneumoniae.
 XX

PN US2002061545-A1.
 XX

PD 23-MAY-2002.
 XX

PF 22-JAN-2001; 2001US-00765272.
 XX

PR 30-OCT-1997; 97US-00961083.
 XX

PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX

DR WPI; 2002-479261/51.
 DR N-PSDB; ABQ84870.
 XX

XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX

PS Claim 11; Page 42; 70pp; English.
 XX

XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX

SQ Sequence 471 AA;
 Query Match 9.2%; Score 76; DB 5; Length 471;
 Best Local Similarity 25.2%; Pred. No. 11;

Matches 30; Conservative 19; Mismatches 48; Indels 22; Gaps 6;
 QY 34 GDLVHERKPGQEPFYVYVIMPKGTEDDQWILESVGGDHYKLGK-----NKFSGRYL 86
 DB 156 GAMQNEWLYDPAYGAYFLKSDGT-YANQEW--QKVGKWIYFKKGMARNEWQG--- 209

QY 87 VYGTDFYLTAGAAVREMDHFKFTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYF 145
 DB 210 -----NYLTGSGA---MATDEVIMDGTTRYIFAASGELKEKDLNVGV-VHRDGRKRYF 259

RESULT 12
 ADC45239
 ID ADC45239 standard; protein; 471 AA.
 XX
 AC ADC45239;
 XX

DT 18-DEC-2003 (first entry)
 XX

DE S. pneumoniae antigenic protein SP091.
 XX

KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
 XX

OS Streptococcus pneumoniae.
 XX

PN US6573082-B1.
 XX

PD 03-JUN-2003.
 XX

PF 28-MAR-2000; 2000US-00536784.
 XX

PR 31-OCT-1996; 96US-0029960P.
 PR 30-OCT-1997; 97US-00961083.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX

DR WPI; 2003-764574/72.
 DR N-PSDB; ADC45238.
 XX

PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 158; 58pp; English.

XX CC The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence
 CC represents an S. pneumoniae antigenic protein.

XX SQ Sequence 471 AA;

Query Match 9.2%; Score 76; DB 7; Length 471;
 Best Local Similarity 25.2%; Pred. No. 11;
 Matches 30; Conservative 19; Mismatches 48; Indels 22; Gaps 6;

Qy 34 GDLVHHERKPGQEPFYVYVIMPKGTEYDDQWILESGDHYKLGK-----NKFSGYL 86
 Db 156 GAMQNEWLYDPAYSAFYLKSDGT-YANQEW--QKVGKWKYFKWGYMARNEWQG--- 209
 Qy 87 VYGFDFYFLTAGAAVRMDHFKFTADGTGKYDIDSSKANGHPRSRGKNWGMKDGKHYF 145
 Db 210 -----NYLITGSGA--MATDEVIMDGTIFYFAASGELKEKKDLNVGW-VHRDGRKRYFF 259

RESULT 13

AA581745
 ID AAY81745 standard; protein; 490 AA.

XX AC AAY81745;

XX DT 02-JUN-2000 (first entry)

XX DE Streptococcus pneumoniae protein sequence ID51.

XX KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.

XX OS Streptococcus pneumoniae.

XX FN WO200006738-A2.

XX PD 10-FEB-2000.

XX PF 27-JUL-1999; 95WO-GB002452.

XX PR 27-JUL-1998; 98GB-00016336.

XX PR 19-MAR-1999; 99US-0125329P.

XX PA (MICR-) MICROBIAL TECHNIQS LTD.

XX PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

XX DR WPI; 2000-195301/17.

XX DR N-PSDB; AAZ91841.

XX PT Streptococcal proteins and polynucleotides useful for diagnosis,
 PT treatment and prophylaxis of bacterial infections.

XX PS Claim 2; Page 54; 76pp; English.

XX CC This sequence represents a Streptococcus pneumoniae protein of the

CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
 CC meningitis

XX SQ Sequence 490 AA;

Query Match 9.2%; Score 76; DB 3; Length 490;
 Best Local Similarity 25.2%; Pred. No. 12;
 Matches 30; Conservative 19; Mismatches 48; Indels 22; Gaps 6;

Qy 34 GDLVHHERKPGQEPFYVYVIMPKGTEYDDQWILESGDHYKLGK-----NKFSGYL 86
 Db 175 GAMQNEWLYDPAYSAFYLKSDGT-YANQEW--QKVGKWKYFKWGYMARNEWQG--- 228

Qy 87 VYGFDFYFLTAGAAVRMDHFKFTADGTGKYDIDSSKANGHPRSRGKNWGMKDGKHYF 145

Db 229 -----NYLITGSGA--MATDEVIMDGTIFYFAASGELKEKKDLNVGW-VHRDGRKRYFF 278

RESULT 14

ABU02034
 ID ABU02034 standard; protein; 506 AA.

XX AC ABU02034;

XX DT 23-OCT-2003 (revised)

XX DT 11-FEB-2003 (first entry)

XX DE S. pneumoniae type 4 strain protein from coding region #1611.

XX KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.

XX OS Streptococcus pneumoniae; type 4 strain.

XX FN WO200277021-A2.

XX PD 03-OCT-2002.

XX PF 27-MAR-2002; 2002WO-IB002163.

XX PR 27-MAR-2001; 2001GB-00007658.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Massignani V, Tettelin H, Fraser C;

XX DR WPI; 2003-040579/03.

XX DR N-PSDB; ABX07323.

XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.

XX PS Claim 1; SEQ ID NO 3222; 56pp; English.

XX CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:29:25 ; Search time 128 Seconds
(without alignments)
380.819 Million cell updates/sec

Title: US-10-768-874-2
Perfect score: 825
Sequence: 1 MSITQSFVLTALFAGASD.....NWGMKGKHYFTVENCQE 152

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	825	100.0	152	14	US-10-218-699-2
2	825	100.0	152	14	US-10-288-740-2
3	825	100.0	152	16	US-10-768-874-2
4	610	73.9	109	14	US-10-218-699-5
5	610	73.9	109	14	US-10-288-740-5
6	610	73.9	109	16	US-10-768-874-5
7	130	15.8	22	14	US-10-218-699-3
8	130	15.8	22	14	US-10-288-740-3
9	130	15.8	22	16	US-10-768-874-3
10	87	10.5	15	14	US-10-218-699-4
11	87	10.5	15	14	US-10-288-740-4
12	87	10.5	15	16	US-10-768-874-4
13	80.5	9.8	280	16	US-10-780-396-17
14	80.5	9.8	329	16	US-10-790-396-7
15	78.5	9.5	721	16	US-10-437-963-193032

16	78	9.5	1143	15	US-10-369-493-1499	Sequence 1499, Ap
17	76.5	9.3	332	12	US-10-440-503-244	Sequence 244, App
18	76.5	9.3	332	12	US-10-461-525-244	Sequence 244, App
19	76.5	9.3	332	14	US-10-146-772-244	Sequence 244, App
20	76.5	9.3	332	15	US-10-241-742-244	Sequence 244, App
21	76.5	9.3	332	15	US-10-440-523-244	Sequence 244, App
22	76	9.2	471	9	US-09-765-272-158	Sequence 158, App
23	76	9.2	490	10	US-09-769-744A-98	Sequence 98, Appl
24	75	9.1	424	16	US-10-229-148B-13	Sequence 13, Appl
25	75	9.1	446	9	US-09-870-472-3	Sequence 3, Appli
26	75	9.1	446	9	US-09-870-472-6	Sequence 6, Appli
27	75	9.1	676	12	US-10-282-122A-74789	Sequence 74789, A
28	75	9.1	983	16	US-10-408-765A-2112	Sequence 2112, Ap
29	75	9.1	1013	14	US-10-205-194-15	Sequence 15, Appl
30	75	9.1	1013	16	US-10-408-765A-379	Sequence 379, App
31	75	9.1	1013	16	US-10-408-765A-2113	Sequence 2113, Ap
32	75	9.1	1015	16	US-10-408-765A-206	Sequence 206, App
33	75	9.1	1015	16	US-10-408-765A-207	Sequence 207, App
34	74.5	9.0	878	12	US-10-282-122A-65010	Sequence 65010, A
35	74.5	9.0	1471	16	US-10-437-963-190303	Sequence 190303, A
36	73	8.8	572	14	US-10-156-761-10627	Sequence 10627, A
37	73	8.8	996	16	US-10-437-963-175079	Sequence 175079, A
38	73	8.8	1034	16	US-10-437-963-175053	Sequence 175053, A
39	72.5	8.8	343	12	US-10-282-122A-51283	Sequence 51283, A
40	72.5	8.8	428	12	US-10-282-122A-46513	Sequence 46513, A
41	72.5	8.8	911	14	US-10-043-344-107	Sequence 107, App
42	72	8.7	255	10	US-09-880-748-1849	Sequence 1849, Ap
43	72	8.7	255	12	US-10-293-418-1849	Sequence 1849, Ap
44	72	8.7	943	9	US-09-969-362-5	Sequence 5, Appli
45	71.5	8.7	358	12	US-10-282-122A-71100	Sequence 71100, A

ALIGNMENTS

RESULT 1
US-10-218-699-2
; Sequence 2, Application US/10218699
; Publication No. US20030012795A1
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
; APPLICANT: Ribeiro, Jose M.C.
; APPLICANT: Cupp, Eddie W.
; APPLICANT: Swaim, Steven F.
; TITLE OF INVENTION: Recombinant Vasoactive Protein from
; TITLE OF INVENTION: Salivary Gland of the Black Fly
; FILE REFERENCE: 35721/205407
; CURRENT APPLICATION NUMBER: US/10/218,699
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: US/09/702,647
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 152
; TYPE: PPT
; ORGANISM: Simulium vittatum
US-10-218-699-2

Query Match	100.0%	Score 825;	DB 14;	Length 152;
Best Local Similarity	100.0%;	Pred. No. 1.8e-86;		
Matches 152;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSITQSFVLTALFAGASDNPIDRKCIVISDGLVMHKKPQGFPPYVYVIMPKGTEY	60	
Db	1	MSITQSFVLTALFAGASDNPIDRKCIVISDGLVMHKKPQGFPPYVYVIMPKGTEY	60	
QY	61	DDQWILSVGGDHYKLNKFSGYLYVGTDFYLTAGAAVREMDHFFKFTADGCKYDIS	120	
Db	61	DDQWILSVGGDHYKLNKFSGYLYVGTDFYLTAGAAVREMDHFFKFTADGCKYDIS	120	
QY	121	SKANGHPSRCKNGVMKDGSKHYFTVENCQE	152	

Db 121 SKANGHPRSRGKNWGMKDGEKHYFTVENCQE 152

RESULT 2

US-10-288-740-2

; Sequence 2, Application US/10288740

; Publication No. US20030109447A1

; GENERAL INFORMATION:

; APPLICANT: Cupp, Mary S.

; Ribeiro, Jose M.C.

; Cupp, Eddie W.

; Swaim, Steven F.

; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM

; SALIVARY GLAND OF THE BLACK FLY

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)

; STREET: 3605 Glenwood Ave. Suite 310

; CITY: Raleigh

; STATE: NC

; COUNTRY: USA

; ZIP: 27622

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/288,740

; FILING DATE: 06-NO. US20030109447A1-2002

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/036,355A

; FILING DATE: 06-MARCH-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Spruill, W. Murray

; REGISTRATION NUMBER: 32,943

; REFERENCE/DOCKET NUMBER: 5721-4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919 420 2202

; TELEFAX: 919 881 3175

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 152 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-288-740-2

Query Match 100.0%; Score 825; DB 14; Length 152;

Best Local Similarity 100.0%; Pred. No. 1.8e-86;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSITQSFVLTALFCAASDNPIADRKCVISDGLVWHERKPGQFPYVYVIMPKGTEY 60

Db 1 MSITQSFVLTALFCAASDNPIADRKCVISDGLVWHERKPGQFPYVYVIMPKGTEY 60

QY 61 DDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAVREMDHFKFTADGTGKYDIS 120

Db 61 DDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAVREMDHFKFTADGTGKYDIS 120

QY 121 SKANGHPRSRGKNWGMKDGEKHYFTVENCQE 152

Db 121 SKANGHPRSRGKNWGMKDGEKHYFTVENCQE 152

RESULT 3

US-10-768-874-2

; Sequence 2, Application US/10768874

; Publication No. US20040142869A1

; GENERAL INFORMATION:

; APPLICANT: Cupp, Mary S.

; APPLICANT: Ribeiro, Jose M.C.

; Cupp, Eddie W.

; APPLICANT: Swaim, Steven F.

; TITLE OF INVENTION: Recombinant Vasoactive Protein from

; Salivary Gland of the Black Fly

; FILE REFERENCE: 35721/205407

; CURRENT APPLICATION NUMBER: US/10/768,874

; CURRENT FILING DATE: 2004-01-30

; PRIOR APPLICATION NUMBER: US/10/218,699

; PRIOR FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: US/09/702,647

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 152

; TYPE: PRT

; ORGANISM: Simulium vittatum

US-10-768-874-2

Query Match 100.0%; Score 825; DB 16; Length 152;

Best Local Similarity 100.0%; Pred. No. 1.8e-86;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSITQSFVLTALFCAASDNPIADRKCVISDGLVWHERKPGQFPYVYVIMPKGTEY 60

Db 1 MSITQSFVLTALFCAASDNPIADRKCVISDGLVWHERKPGQFPYVYVIMPKGTEY 60

QY 61 DDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAVREMDHFKFTADGTGKYDIS 120

Db 61 DDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAVREMDHFKFTADGTGKYDIS 120

QY 121 SKANGHPRSRGKNWGMKDGEKHYFTVENCQE 152

Db 121 SKANGHPRSRGKNWGMKDGEKHYFTVENCQE 152

RESULT 4

US-10-218-699-5

; Sequence 5, Application US/10218699

; Publication No. US20030012795A1

; GENERAL INFORMATION:

; APPLICANT: Cupp, Mary S.

; Ribeiro, Jose M.C.

; Cupp, Eddie W.

; APPLICANT: Swaim, Steven F.

; TITLE OF INVENTION: Recombinant Vasoactive Protein from

; Salivary Gland of the Black Fly

; FILE REFERENCE: 35721/205407

; CURRENT APPLICATION NUMBER: US/10/218,699

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: US/09/702,647

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 109

; TYPE: PRT

; ORGANISM: Simulium vittatum

US-10-218-699-5

Query Match 73.9%; Score 610; DB 14; Length 109;

Best Local Similarity 100.0%; Pred. No. 5.8e-62;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 KPGQFPYVYVIMPKGTEYDDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAV 101

Db 1 KPGQFPYVYVIMPKGTEYDDQRWILESVGGDHYKLNKFGSRILVYGTFFLTAGAAV 60

QY 102 REMDHFKFTADGTGKYDISSKANGHPRSRGKNWGMKDGEKHYFTVENC 150

Db 61 REMDHFKFTADGTGKYDISSKANGHPRSRGKNWGMKDGEKHYFTVENC 109

```
RESULT 5
US-10-288-740-5
; Sequence 5, Application US/10288740
; Publication No. US20030109447A1
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
;               Ribeiro, Jose M.C.
;               Cupp, Eddie W.
;               Swaim, Steven F.
; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
;                   SALIVARY GLAND OF THE BLACK FLY
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/288,740
; FILING DATE: 06-NO. US20030109447A1-2002
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,355A
; FILING DATE: 06-MARCH-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/POCKET NUMBER: 5721-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 5:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 109 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: peptide
;     SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-288-740-5
Query Match          73.9%; Score 610; DB 14; Length 109;
Best Local Similarity 100.0%; Pred. No. 5.8e-62;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 42 KPGQFFPVYVMI PKGTEYDDQRMILSVGGDHYKLNKFKSGRYLVYGTDFYFLTAGAAV 101
Db 1 KPGQFFPVYVMI PKGTEYDDQRMILSVGGDHYKLNKFKSGRYLVYGTDFYFLTAGAAV 60
Qy 102 REMDHFXTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYFTVENC 150
Db 61 REMDHFXTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYFTVENC 109
RESULT 6
US-10-768-874-5
; Sequence 5, Application US/10768874
; Publication No. US20040142869A1
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
;               Ribeiro, Jose M.C.
;               Cupp, Eddie W.
;               Swaim, Steven F.
; TITLE OF INVENTION: Recombinant Vasoactive Protein from
;                   Salivary Gland of the Black Fly
; NUMBER OF SEQUENCES: 5
```

```
FILE REFERENCE: 35721/205407
CURRENT APPLICATION NUMBER: US/10/768,874
CURRENT FILING DATE: 2004-01-30
PRIOR APPLICATION NUMBER: US/10/218,699
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: US/09/702,647
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 109
TYPE: PRT
ORGANISM: Simulium vittatum
US-10-768-874-5
Query Match          73.9%; Score 610; DB 16; Length 109;
Best Local Similarity 100.0%; Pred. No. 5.8e-62;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 42 KPGQFFPVYVMI PKGTEYDDQRMILSVGGDHYKLNKFKSGRYLVYGTDFYFLTAGAAV 101
Db 1 KPGQFFPVYVMI PKGTEYDDQRMILSVGGDHYKLNKFKSGRYLVYGTDFYFLTAGAAV 60
Qy 102 REMDHFXTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYFTVENC 150
Db 61 REMDHFXTADGTGKYDISSKANGHPRSRGKNWGMKDGKHYFTVENC 109
RESULT 7
US-10-218-699-3
; Sequence 3, Application US/10218699
; Publication No. US20030012795A1
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
;               Ribeiro, Jose M.C.
;               Cupp, Eddie W.
;               Swaim, Steven F.
; TITLE OF INVENTION: Recombinant Vasoactive Protein from
;                   Salivary Gland of the Black Fly
; FILE REFERENCE: 35721/205407
; CURRENT APPLICATION NUMBER: US/10/218,699
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: US/09/702,647
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Simulium vittatum
US-10-218-699-3
Query Match          15.8%; Score 130; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 131 GKNWGMKDGKHYFTVENCQE 152
Db 1 GKNWGMKDGKHYFTVENCQE 22
RESULT 8
US-10-288-740-3
; Sequence 3, Application US/10288740
; Publication No. US20030109447A1
; GENERAL INFORMATION:
; APPLICANT: Cupp, Mary S.
;               Ribeiro, Jose M.C.
;               Cupp, Eddie W.
;               Swaim, Steven F.
; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
;                   SALIVARY GLAND OF THE BLACK FLY
; NUMBER OF SEQUENCES: 5
```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
;; STREET: 3605 Glenwood Ave. Suite 310
;; CITY: Raleigh
;; STATE: NC
;; COUNTRY: USA
;; ZIP: 27622
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/288,740
;; FILING DATE: 06-NO. US20030109447A1-2002
;; CLASSIFICATION: 514
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/09/036,355A
;; FILING DATE: 06-MARCH-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Spruill, W. Murray
;; REGISTRATION NUMBER: 32,943
;; REFERENCE/DOCKET NUMBER: 5721-4
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919 420 2202
;; TELEFAX: 919 881 3175
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 22 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-288-740-3

Query Match 15.8%; Score 130; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 GKNWGVMDGKGYFTVENCQ 152
DB 1 GKNWGVMDGKGYFTVENCQ 22

RESULT 9
US-10-768-874-3
;; Sequence 3, Application US/10768874
;; Publication No. US20040142869A1
;; GENERAL INFORMATION:
;; APPLICANT: Cupp, Mary S.
;; APPLICANT: Ribeiro, Jose M.C.
;; APPLICANT: Cupp, Eddie W.
;; APPLICANT: Swaim, Steven F.
;; TITLE OF INVENTION: Recombinant Vasoactive Protein from
;; TITLE OF INVENTION: Salivary Gland of the Black Fly
;; FILE REFERENCE: 35721/205407
;; CURRENT APPLICATION NUMBER: US/10/768,874
;; CURRENT FILING DATE: 2004-01-30
;; PRIOR APPLICATION NUMBER: US/10/218,699
;; PRIOR FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: US/09/702,647
;; PRIOR FILING DATE: 2000-10-31
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 22
;; TYPE: PRT
;; ORGANISM: Simulium vittatum
US-10-768-874-3

Query Match 15.8%; Score 130; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e-08;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 131 GKNWGVMDGKGYFTVENCQ 152
DB 1 GKNWGVMDGKGYFTVENCQ 22

RESULT 10
US-10-218-699-4
;; Sequence 4, Application US/10218699
;; Publication No. US20030012795A1
;; GENERAL INFORMATION:
;; APPLICANT: Cupp, Mary S.
;; APPLICANT: Ribeiro, Jose M.C.
;; APPLICANT: Cupp, Eddie W.
;; APPLICANT: Swaim, Steven F.
;; TITLE OF INVENTION: Recombinant Vasoactive Protein from
;; TITLE OF INVENTION: Salivary Gland of the Black Fly
;; FILE REFERENCE: 35721/205407
;; CURRENT APPLICATION NUMBER: US/10/218,699
;; CURRENT FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: US/09/702,647
;; PRIOR FILING DATE: 2000-10-31
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4
;; LENGTH: 15
;; TYPE: PRT
;; ORGANISM: Simulium vittatum
US-10-218-699-4

Query Match 10.5%; Score 87; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 KPGQFFPYVYMIPK 56
DB 1 KPGQFFPYVYMIPK 15

RESULT 11
US-10-288-740-4
;; Sequence 4, Application US/10288740
;; Publication No. US20030109447A1
;; GENERAL INFORMATION:
;; APPLICANT: Cupp, Mary S.
;; APPLICANT: Ribeiro, Jose M.C.
;; APPLICANT: Cupp, Eddie W.
;; APPLICANT: Swaim, Steven F.
;; TITLE OF INVENTION: RECOMBINANT VASOACTIVE PROTEIN FROM
;; TITLE OF INVENTION: SALIVARY GLAND OF THE BLACK FLY
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
;; STREET: 3605 Glenwood Ave. Suite 310
;; CITY: Raleigh
;; STATE: NC
;; COUNTRY: USA
;; ZIP: 27622
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/288,740
;; FILING DATE: 06-NO. US20030109447A1-2002
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/09/036,355A
;; FILING DATE: 06-MARCH-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Spruill, W. Murray

REGISTRATION NUMBER: 32,943
 REFERENCE/DOCKET NUMBER: 5721-4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919 420 2202
 TELEFAX: 919 881 3175
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 15 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-10-288-740-4
 Query Match 10.5%; Score 87; DB 14; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0052;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 KPGQEFPPYVYMIPK 56
 Db 1 KPGQEFPPYVYMIPK 15

RESULT 12

US-10-768-874-4

Sequence 4, Application US/10768874
 Publication No. US20040142869A1

GENERAL INFORMATION:
 APPLICANT: Cupp, Mary S.
 APPLICANT: Ribeiro, Jose M.C.
 APPLICANT: Cupp, Eddie W.
 APPLICANT: Swaim, Steven F.
 TITLE OF INVENTION: Recombinant Vasoactive Protein from
 TITLE OF INVENTION: Salivary Gland of the Black Fly
 FILE REFERENCE: 35721/205407
 CURRENT APPLICATION NUMBER: US/10/768,874
 CURRENT FILING DATE: 2004-01-30
 PRIOR APPLICATION NUMBER: US/10/218,699
 PRIOR FILING DATE: 2002-08-14
 PRIOR APPLICATION NUMBER: US/09/702,647
 PRIOR FILING DATE: 2000-10-31
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 4
 LENGTH: 15
 TYPE: PRT
 ORGANISM: Simulium vittatum

US-10-768-874-4
 Query Match 10.5%; Score 87; DB 16; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0052;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 KPGQEFPPYVYMIPK 56
 Db 1 KPGQEFPPYVYMIPK 15

RESULT 13

US-10-396-17

Sequence 17, Application US/10790396
 Publication No. US20040157296A1

GENERAL INFORMATION:
 APPLICANT: Sim, Gek-kee
 APPLICANT: Yang, Shumin
 APPLICANT: Sellins, Karen S.
 TITLE OF INVENTION: NOVEL FORMS OF T CELL COSTIMULATORY
 TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
 FILE REFERENCE: IM-1-CL-PCT
 CURRENT APPLICATION NUMBER: US/10/790,396
 CURRENT FILING DATE: 2004-03-01
 PRIOR APPLICATION NUMBER: US/09/646,561

PRIOR FILING DATE: 2000-09-19
 PRIOR APPLICATION NUMBER: 60/078,765
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 09/062,597
 PRIOR FILING DATE: 1998-04-17
 NUMBER OF SEQ ID NOS: 65
 SOFTWARE: Patent in Ver. 2.0
 SEQ ID NO 17

LENGTH: 280
 TYPE: PRT
 ORGANISM: Canis familiaris
 US-10-790-396-17

Query Match 9.8%; Score 80.5; DB 16; Length 280;
 Best Local Similarity 27.1%; Pred. No. 1.3;
 Matches 29; Conservative 13; Mismatches 38; Indels 27; Gaps 4;
 Qy 1 MSITQSPFVLTALIFGAA-----SDNPIADRKCVISDGD-LVMH 39
 Db 7 MELANNILFVMTLLYGAASMSQAYFNKTGELPCHFTNSQNISLDELVWFWDQDKLVLY 66
 Qy 40 ERKPGQEFPPYVYMIPKG-TEYDDQRMWILESGDGHYKLNKPKSGRY 85
 Db 67 ELYRGKENPQNVHRKYGRTSFDKNWTLRL-----HNIQIKDKGLY 108

RESULT 14

US-10-790-396-7

Sequence 7, Application US/10790396
 Publication No. US20040157296A1

GENERAL INFORMATION:
 APPLICANT: Sim, Gek-kee
 APPLICANT: Yang, Shumin
 APPLICANT: Sellins, Karen S.
 TITLE OF INVENTION: NOVEL FORMS OF T CELL COSTIMULATORY
 TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
 FILE REFERENCE: IM-1-CL-PCT
 CURRENT APPLICATION NUMBER: US/10/790,396
 CURRENT FILING DATE: 2004-03-01
 PRIOR APPLICATION NUMBER: US/09/646,561
 PRIOR FILING DATE: 2000-09-19
 PRIOR APPLICATION NUMBER: 60/078,765
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 09/062,597
 PRIOR FILING DATE: 1998-04-17
 NUMBER OF SEQ ID NOS: 65
 SOFTWARE: Patent in Ver. 2.0
 SEQ ID NO 7
 LENGTH: 329
 TYPE: PRT
 ORGANISM: Canis familiaris

US-10-790-396-7

Query Match 9.8%; Score 80.5; DB 16; Length 329;
 Best Local Similarity 27.1%; Pred. No. 1.6;
 Matches 29; Conservative 13; Mismatches 38; Indels 27; Gaps 4;

Qy 1 MSITQSPFVLTALIFGAA-----SDNPIADRKCVISDGD-LVMH 39
 Db 7 MELANNILFVMTLLYGAASMSQAYFNKTGELPCHFTNSQNISLDELVWFWDQDKLVLY 66
 Qy 40 ERKPGQEFPPYVYMIPKG-TEYDDQRMWILESGDGHYKLNKPKSGRY 85
 Db 67 ELYRGKENPQNVHRKYGRTSFDKNWTLRL-----HNIQIKDKGLY 108

RESULT 15

US-10-437-963-193032

Sequence 193032, Application US/10437963
 Publication No. US20040123343A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.

